STAT 571 — Probability and Measure

Abbreviated Course Notes

Aayush Bajaj
after lectures by Adam B. Kashlak

May 2026

Contents

1 Lecture 1 — Measures and σ-Fields

The opening question of the course is disarmingly simple: what is a measure? Intuitively, a measure assigns a size — length, area, volume, probability — to a set. To pin this down, we have to settle two things up front: which subsets of an ambient space Ω we are even allowed to talk about, and what rules the size-assignment must obey. The first question is answered by a σ-field; the second by the definition of a measure. The relationship of $\mathbb{R}^n$ to its norm sits in the background as the running prototype.

1.1 Notation and set-theoretic preliminaries

Throughout the course Ω denotes a fixed ambient set, the sample space. For $A \subseteq \Omega$ we write

$$A^c \;=\; \{\, x \in \Omega : x \notin A \,\} \;=\; \Omega \setminus A$$

for the complement, and use $\Omega \setminus A$ and A^c interchangeably. A countable collection $\{A_i\}_{i=1}^{\infty}$ of subsets of Ω is pairwise disjoint if

$$A_i \cap A_j \;=\; \varnothing \qquad \text{whenever } i \neq j.$$

The power set of Ω, denoted $\mathcal{P}(\Omega)$, is the collection of all subsets of Ω. When Ω has n elements this collection has 2^n elements, which is the source of the alternative notation 2^{Ω}.

Remark 1.1. Symmetric difference $A \triangle B = (A \setminus B) \cup (B \setminus A)$ is the set-theoretic counterpart of the logical XOR. It will not feature heavily in this lecture, but is worth keeping in the toolkit.

1.2 σ-fields

The first object we need is a class of subsets that is closed under the operations we plan to perform on it: complement, countable union, and (as a consequence) countable intersection.

Definition 1.1: σ-field

For a set Ω, a σ-field (or σ-algebra) is a collection $\mathcal{F}$ of subsets $A \subseteq \Omega$ such that

1. $\varnothing,\ \Omega \in \mathcal{F}$;
2. if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$;
3. for any countable collection $\{A_i\}_{i=1}^{\infty}$ with $A_i \in \mathcal{F}$ for all $i \in \mathbb{N}$,

$$\bigcup_{i=1}^{\infty} A_i \;\in\; \mathcal{F}.$$

The pair $(\Omega, \mathcal{F})$ is then called a measurable space.

Remark 1.2. Combining (2) and (3) via De Morgan's laws delivers closure under countable intersection: for $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$,

$$\Big(\bigcap_{i=1}^{\infty} A_i\Big)^c \;=\; \bigcup_{i=1}^{\infty} A_i^c \;\in\; \mathcal{F},$$

and applying (2) once more yields $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Remark 1.3. Any σ-field on Ω is a subcollection of the power set, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. The power set itself is a σ-field, but it is typically too large to be useful: on $\mathbb{R}$, for example, it is too generous a collection to admit a translation-invariant length-like measure (a fact made precise in Lecture 4).

1.3 Measures

With a class of admissible sets in hand, a measure is a rule for assigning a non-negative size to each.

Definition 1.2: Measure

Let $(\Omega, \mathcal{F})$ be a measurable space. A measure is a function $\mu : \mathcal{F} \to \mathbb{R}^+$ satisfying

1. $\mu(\varnothing) = 0$;
2. μ is countably additive: for any pairwise disjoint countable collection $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

The triple $(\Omega, \mathcal{F}, \mu)$ is then called a measure space.

Remark 1.4. Take care to distinguish a measurable space $(\Omega, \mathcal{F})$ from a measure space $(\Omega, \mathcal{F}, \mu)$: the former specifies only which sets can be sized, the latter also specifies how.

Remark 1.5. There also exist signed measures, which are allowed to take values in $\mathbb{R}$ rather than $\mathbb{R}^+$. These will not concern us in this lecture.

1.4 Special classes of measures

Three size-on-the-whole-space conditions get repeated names.

Definition 1.3: Probability, finite, and σ-finite measures

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

- μ is a probability measure if $\mu(\Omega) = 1$; we then call $(\Omega, \mathcal{F}, \mu)$ a probability space and usually write $\mu = \mathbb{P}$.
- μ is a finite measure if $\mu(\Omega) < \infty$.
- μ is a σ-finite measure if there exists a countable cover $\Omega = \bigcup_{i=1}^{\infty} A_i$ with $A_i \in \mathcal{F}$ and $\mu(A_i) < \infty$ for every i.

■ **Example 1.1 (Length on $\mathbb{R}$).** Take $\Omega = \mathbb{R}$ and (anticipating Lecture 2) define $\mu([a, b]) = b - a$. Then μ is not finite, but it is σ-finite: writing

$$\mathbb{R} = \bigcup_{i=1}^{\infty} ([\, i-1, i\,] \cup [-i, -i+1]),$$

each piece has finite length. This is the prototype of Lebesgue measure.

1.5 Examples on a finite sample space

■ **Example 1.2 (Counting measure).** Let $\Omega = \{1,2,\dots,n\}$ and take $\mathcal{F} = \mathcal{P}(\Omega)$, which has 2^n elements (hence the notation 2^Ω). The counting measure is

$$\mu(A) \;=\; \#A, \qquad A \in \mathcal{F}.$$

Thus $\mu(\{1,3,7\}) = 3$ and $\mu(\Omega) = n$. The normalised version

$$\nu(A) \;=\; \tfrac{1}{n}\,\mu(A)$$

is the uniform probability measure on $\{1,\dots,n\}$.

■ **Example 1.3 (Binomial probability measure).** By contrast, on $\Omega = \{0,1,\dots,n\}$ the binomial (n,p) distribution assigns to each point i the weight

$$\mu(\{i\}) \;=\; \binom{n}{i} p^i\,(1-p)^{\,n-i}, \qquad p \in (0,1),$$

extended additively to $\mathcal{P}(\Omega)$. This gives a probability measure on Ω which is not uniform.

2 Lecture 2 — Constructing σ-Fields and Measures (Existence; Carathéodory's Extension)

Motivating question. Suppose we declare the measure of a half-open interval to be its length, $\mu((a,b]) = b - a$ for $b > a$. What else can we then measure? The collection of all such intervals is not a σ-field — e.g. $(a,b] \cup (c,d]$ is not in general a half-open interval — so we need a procedure that grows the class of "measurable" sets and extends μ to it. The extension is delivered by Carathéodory's extension theorem, the existence half of the construction of a measure space. Uniqueness is taken up next lecture.

2.1 Semirings, rings, and fields

We climb a small ladder of set systems sitting below a σ-field. At each rung the candidate measure has more room to manoeuvre.

Definition 2.1: Semiring

A collection $\mathcal{A}$ of subsets of Ω is a semiring if

- $\varnothing \in \mathcal{A}$,
- $A \cap B \in \mathcal{A}$ for all $A, B \in \mathcal{A}$, and
- for all $A, B \in \mathcal{A}$ the set difference splits as a finite disjoint union

$$B \setminus A = \bigsqcup_{i=1}^{n} C_i, \qquad C_i \in \mathcal{A}.$$

The set difference itself need not lie in $\mathcal{A}$, but it must be expressible as a finite union of members of $\mathcal{A}$.

■ **Example 2.1 (Half-open intervals form a semiring).** The collection of all half-open intervals $(a,b] \subseteq \mathbb{R}$ (together with $\varnothing$) is a semiring: intersections of half-open intervals are half-open intervals, and a difference $(a,b] \setminus (c,d]$ is the union of at most two half-open intervals.

Definition 2.2: Ring

A collection $\mathcal{A}$ of subsets of Ω is a ring if

- $\varnothing \in \mathcal{A}$, and
- for all $A, B \in \mathcal{A}$, both $B \setminus A \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$.

A ring is closed under finite (set-theoretic) unions and differences.

■ **Example 2.2 (Finite unions of half-open intervals).** The collection of all finite unions of half-open intervals $(a,b] \subseteq \mathbb{R}$ is a ring. It is not yet a σ-field — it fails to absorb countable unions.

Definition 2.3: Field

A ring $\mathcal{A}$ is a field (or algebra) on Ω if additionally $\Omega \in \mathcal{A}$.

Remark 2.1. A field that is closed under countable unions is a σ-field. The progression

$$\text{semiring} \subset \text{ring} \subset \text{field} \subset \sigma\text{-field}$$

mirrors a progression in stability under set operations: pairwise intersection only, then finite unions/differences, then Ω, finally countable unions.

2.2 Set functions and pre-measures

Before defining a measure we collect the regularity properties a set function may enjoy.

Definition 2.4: Set function and its properties

Let $\mathcal{A}$ be a collection of subsets of Ω. A set function is any map $\mu\colon \mathcal{A} \to [0,\infty]$ (not necessarily a measure). For $A, B \in \mathcal{A}$ we say:

- μ is increasing (monotone) if $A \subseteq B \implies \mu(A) \le \mu(B)$;
- μ is (finitely) additive if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \mathcal{A}$ are disjoint and $A \cup B \in \mathcal{A}$;
- μ is countably additive if for every pairwise disjoint sequence $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ with $\bigcup_i A_i \in \mathcal{A}$,
$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i);$$
- μ is countably subadditive if for every (not necessarily disjoint) sequence $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ with $\bigcup_i A_i \in \mathcal{A}$,
$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mu(A_i).$$

Definition 2.5: Pre-measure

A set function $\mu\colon \mathcal{A} \to [0,\infty]$ on a ring $\mathcal{A}$ is a pre-measure if $\mu(\varnothing) = 0$ and μ is countably additive on $\mathcal{A}$.

A pre-measure is exactly what a measure looks like before the underlying set system has been closed up to a σ-field. Carathéodory's theorem will perform that closing-up.

2.3 Outer measure and μ^*-measurability

A pre-measure on a ring can be extended in a canonical way to every subset of Ω by approximating from above.

Definition 2.6: Outer measure

Let μ be a pre-measure on a ring $\mathcal{A}$ on Ω. The outer measure induced by μ is

$$\mu^*(E) = \inf\left\{\sum_i \mu(A_i) \;:\; A_i \in \mathcal{A},\ E \subseteq \bigcup_i A_i\right\}, \qquad E \subseteq \Omega,$$

where the infimum runs over finite or countable covers of E by elements of $\mathcal{A}$.

The outer measure μ^* is defined on all of $\mathcal{P}(\Omega)$, but in general it is only countably subadditive there — not additive. To recover countable additivity we restrict attention to sets that "split" every test set cleanly.

Definition 2.7: μ^*-measurable set

A set $B \subseteq \Omega$ is μ^*-measurable (in the sense of Carathéodory) if for every $E \subseteq \Omega$,

$$\mu^*(E \cap B) \; + \; \mu^*(E \cap B^c) \; = \; \mu^*(E).$$

Write $\mathcal{M}$ for the collection of all μ^*-measurable subsets of Ω.

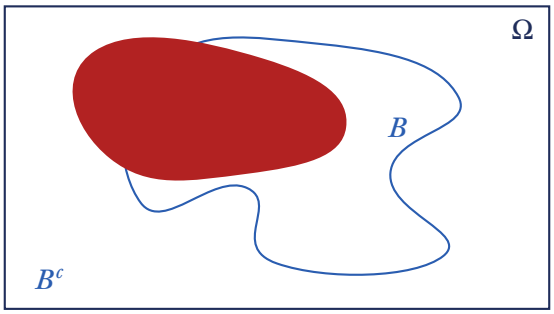


Figure 1. The Carathéodory criterion. A set $B \subseteq \Omega$ is μ^*-measurable when every test set E (red) is split additively by B and its complement: $\mu^*(E) = \mu^*(E \cap B) + \mu^*(E \cap B^c)$.

Remark 2.2. Countable subadditivity of μ^* already gives "$\geq$" in the defining identity, so the substantive condition is the reverse inequality: B does not waste mass at its boundary. The class $\mathcal{M}$ is the largest natural domain on which μ^* is genuinely additive.

2.4 Carathéodory's extension theorem

We now assemble the pieces. Starting from a pre-measure on a ring, the outer measure machinery delivers a measure on a σ-field containing the original ring.

Theorem 2.8: Carathéodory Extension

Let $\mathcal{A}$ be a ring on Ω and let μ be a pre-measure on $\mathcal{A}$. Let μ^* be the outer measure induced by μ and $\mathcal{M}$ the collection of μ^*-measurable sets. Then:

1. $\mu^*(\varnothing) = 0$, and μ^* is monotone and countably subadditive on $\mathcal{P}(\Omega)$;
2. μ^* and μ agree on $\mathcal{A}$, i.e. $\mu^*(A) = \mu(A)$ for every $A \in \mathcal{A}$;
3. $\mathcal{A} \subseteq \mathcal{M}$;
4. $\mathcal{M}$ is a σ-field on Ω and μ^* restricted to $\mathcal{M}$ is a measure;
5. consequently
$$\mathcal{A} \; \subseteq \; \sigma(\mathcal{A}) \; \subseteq \; \mathcal{M} \; \subseteq \; \mathcal{P}(\Omega),$$
and $\mu^*\big|_{\sigma(\mathcal{A})}$ is a measure on $\sigma(\mathcal{A})$ extending μ.

Remark 2.3. The chain in (5) records the precise sense in which Carathéodory "extends" μ: the original ring $\mathcal{A}$ sits inside the generated σ-field $\sigma(\mathcal{A})$, which sits inside the larger σ-field $\mathcal{M}$ of μ^*-measurable sets, which sits inside the full power set. Both $\sigma(\mathcal{A})$ and $\mathcal{M}$ carry the measure μ^*;

the "correct" extension is the outer measure restricted to $\sigma(\mathcal{A})$. It can happen that $\sigma(\mathcal{A})$ is a strict subset of $\mathcal{M}$ (this is the gap filled by completion).

Remark 2.4. We have shown existence: at least one measure on $\sigma(\mathcal{A})$ agreeing with μ on $\mathcal{A}$ exists. The companion question — is the extension unique? — is

$$\mu_1(A) = \mu_2(A)\ \forall A \in \mathcal{A} \quad \overset{?}{\Longrightarrow} \quad \mu_1(B) = \mu_2(B)\ \forall B \in \sigma(\mathcal{A}),$$

and is the subject of the next lecture, via Dynkin's π-λ theorem.

3 Lecture 3 — Uniqueness; Dynkin's π–λ Theorem; Completeness

Lecture 2 produced an extension of a pre-measure to the generated σ-field via Carathéodory. The natural follow-up is the question that opens the handwritten page:

> If $\mu_1(A) = \mu_2(A)$ for every $A \in \mathcal{A}$, does it follow that $\mu_1(B) = \mu_2(B)$ for every $B \in \sigma(\mathcal{A})$?

The answer is yes provided $\mathcal{A}$ is a π-system and the measures are σ-finite. The vehicle is Dynkin's π–λ theorem, which we develop next, after which we take a brief detour through completeness.

3.1 π- and λ-systems

Definition 3.1: π-system

A collection $\mathcal{A}$ of subsets of Ω is a π-system if it is closed under finite intersections: for every $A, B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$.

Definition 3.2: λ-system

A collection $\mathcal{L}$ of subsets of Ω is a λ-system if

- $\Omega \in \mathcal{L}$;
- for any $A, B \in \mathcal{L}$ with $A \subset B$, $B \setminus A \in \mathcal{L}$;
- for any pairwise-disjoint sequence $\{A_i\}_{i=1}^{\infty} \subset \mathcal{L}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$.

Remark 3.1. A λ-system looks much like a σ-field, the difference being that countable unions are required only for disjoint sequences. Every field is automatically a π-system, and a collection that is both a π-system and a λ-system is a σ-field.

3.2 Dynkin's π–λ theorem

The lecture's central tool upgrades containment in a λ-system to containment of the full generated σ-field, provided one starts from a π-system.

Theorem 3.3: Dynkin π–λ theorem

Let $\mathcal{A}$ be a π-system and $\mathcal{L}$ a λ-system on Ω with $\mathcal{A} \subset \mathcal{L}$. Then $\sigma(\mathcal{A}) \subset \mathcal{L}$.

Remark 3.2. The proof strategy from the handwriting: take $\mathcal{L}_0$ to be the smallest λ-system containing $\mathcal{A}$. Show $\mathcal{L}_0$ is also closed under intersections, by introducing

$$\mathcal{L}' = \{B \in \mathcal{L}_0 : B \cap A \in \mathcal{L}_0 \text{ for all } A \in \mathcal{A}\}$$

and verifying $\mathcal{L}'$ is a λ-system containing $\mathcal{A}$, hence $\mathcal{L}' = \mathcal{L}_0$; then repeat the argument with $\mathcal{L}'' = \{B \in \mathcal{L}_0 : B \cap C \in \mathcal{L}_0 \text{ for all } C \in \mathcal{L}_0\}$. A λ-system that is also a π-system is a σ-field, so $\sigma(\mathcal{A}) \subseteq \mathcal{L}_0 \subseteq \mathcal{L}$.

3.3 Uniqueness of extension

The promised payoff:

Theorem 3.4: Uniqueness of extension

Let $\mathcal{A}$ be a π-system on Ω and let μ_1, μ_2 be two σ-finite measures on $\sigma(\mathcal{A})$. If $\mu_1(A) = \mu_2(A)$ for every $A \in \mathcal{A}$, then $\mu_1(B) = \mu_2(B)$ for every $B \in \sigma(\mathcal{A})$.

Remark 3.3. The handwriting splits the proof in two:

- Finite case. Assume $\mu_1(\Omega) = \mu_2(\Omega) < \infty$ and set $\mathcal{L} = \{B \subset \Omega : \mu_1(B) = \mu_2(B)\}$. One verifies $\mathcal{L}$ is a λ-system: $\Omega \in \mathcal{L}$ by hypothesis; if $A \subset B$ lie in $\mathcal{L}$ then $\mu_i(B \setminus A) = \mu_i(B) - \mu_i(A)$ (here finiteness allows the subtraction), so $B \setminus A \in \mathcal{L}$; countable additivity handles disjoint unions. Since $\mathcal{A} \subset \mathcal{L}$, Dynkin gives $\sigma(\mathcal{A}) \subset \mathcal{L}$.
- σ-finite case. For each $A \in \mathcal{A}$ with $\mu_1(A) = \mu_2(A) < \infty$, define $\mathcal{L}_A = \{B \subseteq \Omega : \mu_1(A \cap B) = \mu_2(A \cap B)\}$, a λ-system; by Dynkin $\sigma(\mathcal{A}) \subset \mathcal{L}_A$. Decompose $\Omega = \bigcup_{i \geq 1} A_i$ with $A_i \in \mathcal{A}$ and $\mu_1(A_i) = \mu_2(A_i) < \infty$. Inclusion–exclusion on the finitely many $A_1, \ldots, A_n$ (using that $\mathcal{A}$ is a π-system, so $A_i \cap A_j \in \mathcal{A}$ and so on) gives $\mu_1(B \cap \bigcup_{i \leq n} A_i) = \mu_2(B \cap \bigcup_{i \leq n} A_i)$ for every $B \in \sigma(\mathcal{A})$; let $n \to \infty$.

Remark 3.4. π-systems are very natural in probability theory: the joint event $A \cap B$ lives alongside A and B, so any sensible event class is closed under finite intersections.

■ **Example 3.1 (σ-finiteness is needed).** Take $\Omega = (0,1]$ and let $\mathcal{A}$ be the collection of finite unions of half-open intervals $(a,b]$. Two measures on $\sigma(\mathcal{A})$ agreeing on $\mathcal{A}$ but disagreeing on $\sigma(\mathcal{A})$:

- μ sends $\varnothing \mapsto 0$ and every non-empty element of $\mathcal{A}$ to ∞; the induced outer measure μ^* then assigns ∞ to every non-empty subset of Ω.
- Counting measure ν also sends $\varnothing \mapsto 0$ and each non-empty $(a,b]$ to ∞, but $\nu(\{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}) = 3$ while $\mu^*(\{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}) = \infty$.

The pre-measure on $\mathcal{A}$ is not σ-finite, and uniqueness breaks at the level of $\sigma(\mathcal{A})$.

3.4 Completeness

We now leave uniqueness behind and address a different deficiency: a measure space may contain sets of measure zero whose subsets are not themselves measurable. We patch this by enlarging the σ-field to absorb every such "negligible" set. The geometry behind the patch is the symmetric difference.

Definition 3.5: Symmetric difference

For sets $A, B \subseteq \Omega$,

$$A \triangle B \;=\; (A \setminus B) \cup (B \setminus A).$$

For a measure space $(X, \mathcal{F}, \mu)$ the outer measure extends μ to all of $\mathcal{P}(X)$ by

$$\mu^*(B) \;=\; \inf\{\, \mu(A) : B \subset A,\ A \in \mathcal{F} \,\}.$$

A subset $N \subseteq X$ is μ-null if $\mu^*(N) = 0$; write $\mathcal{N}_\mu$ for the collection of all such sets.

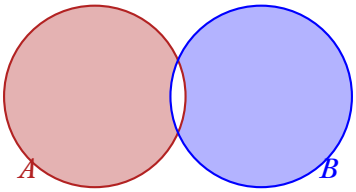


Figure 2. The shaded region is $A \triangle B$; the unshaded overlap is $A \cap B$.

Definition 3.6: Complete measure space

The measure space $(X, \mathcal{F}, \mu)$ is complete if every μ-null set already lies in $\mathcal{F}$, i.e. $\mathcal{N}_\mu \subset \mathcal{F}$.

Remark 3.5 (Exercise from the lecture)**.** Show that $\mathcal{N}_\mu$ is a ring of sets.

When $(X, \mathcal{F}, \mu)$ fails to be complete, we can replace $\mathcal{F}$ by a slightly larger σ-field that contains all null sets.

Proposition 3.7: Completion (Dudley 3.3.2)

Let $(X, \mathcal{F}, \mu)$ be a measure space with null-set collection $\mathcal{N}_\mu$. Define

$$\mathcal{F} \vee \mathcal{N}_\mu \;=\; \{\, A \cup N : A \in \mathcal{F},\ N \in \mathcal{N}_\mu \,\}.$$

Then

$$\mathcal{F} \vee \mathcal{N}_\mu \;=\; \{\, B \subseteq X : \exists\, A \in \mathcal{F} \text{ with } A \triangle B \in \mathcal{N}_\mu \,\},$$

and this is the smallest σ-field containing both $\mathcal{F}$ and $\mathcal{N}_\mu$. Setting $\bar{\mu}(A \cup N) = \mu(A)$, the triple $(X, \mathcal{F} \vee \mathcal{N}_\mu, \bar{\mu})$ is a complete measure space, the completion of $(X, \mathcal{F}, \mu)$.

Remark 3.6. The two descriptions of $\mathcal{F} \vee \mathcal{N}_\mu$ match the picture above: B is in the completion iff it differs from some $A \in \mathcal{F}$ only on a set of outer measure zero, i.e. $A \triangle B$ is null. Lebesgue measure on $\mathbb{R}$ is the completion of its restriction to the Borel σ-field; this strict enlargement is the gap $\mathcal{B} \subsetneq \mathcal{M}_\lambda$ used in the next lecture.

4 Lecture 4 — Lebesgue Measure; Non-Measurable Sets; Product Measures and Independence

The Carathéodory and Dynkin machinery of Lecture 3 is now put to work. We build Lebesgue measure on $(0,1]$ (and on $\mathbb{R}$), exhibit a Vitali set that no translation-invariant measure can size, then sketch how the same toolkit yields product measures and independence of σ-fields.

4.1 Lebesgue measure on $(0,1]$ and $\mathbb{R}$

Take $\Omega = (0,1]$ (or $\Omega = \mathbb{R}$) and let $\mathcal{A}$ be the collection of all finite unions of half-open intervals $(a,b]$, together with $\varnothing$. The target is a set function with

$$\lambda\big((a,b]\big) \;=\; b-a,$$

extended by additivity to $\mathcal{A}$.

Proposition 4.1: $\mathcal{A}$ is a π-system and a ring

The collection $\mathcal{A}$ of finite (disjoint) unions of half-open intervals $(a,b] \subseteq (0,1]$, together with $\varnothing$, is a π-system: intersections of half-open intervals are again half-open intervals,

$$(a,b]\cap(c,d] \;=\; \begin{cases} \varnothing & \text{if } b \le c, \\ (c,b] & \text{if } a \le c < b \le d, \end{cases}$$

and unions of finitely many such intersections remain in $\mathcal{A}$. The class $\mathcal{A}$ is also a ring: it contains $\varnothing = (a,a]$, is closed under set difference $(a,b]\setminus(c,d]$, and closed under finite unions by definition.

Proposition 4.2: λ is a pre-measure on $\mathcal{A}$

The set function $\lambda\colon \mathcal{A} \to [0,\infty]$ defined by $\lambda((a,b]) = b-a$ and extended additively to finite disjoint unions is a pre-measure on the ring $\mathcal{A}$.

Combining the two propositions with the Carathéodory and π-λ theorems of Lecture 3 produces the construction.

Theorem 4.3: Lebesgue measure

Let $\mathcal{B} = \sigma(\mathcal{A})$ be the Borel σ-field of $(0,1]$ (equivalently, the σ-field generated by the open sets). There exists a unique measure λ on $\mathcal{B}$ with $\lambda((a,b]) = b-a$. It is obtained by Carathéodory extension (Step 2) of the pre-measure on $\mathcal{A}$ (Step 1); uniqueness on $\sigma(\mathcal{A})$ follows because $\mathcal{A}$ is a π-system (Step 3). Writing $\mathcal{M}_\lambda$ for the σ-field of Lebesgue measurable sets,

$$\mathcal{A} \;\subset\; \mathcal{B} \;=\; \sigma(\mathcal{A}) \;\subsetneq\; \mathcal{M}_\lambda \;\subsetneq\; \mathcal{P}\big((0,1]\big),$$

where $\mathcal{M}_\lambda$ is the completion of $\mathcal{B}$ with respect to the λ-null sets $\mathcal{N}_\lambda$.

Remark 4.1. Lebesgue measure is the only translation-invariant measure on $\mathbb{R}$ (up to a multiplicative constant) assigning length $b-a$ to $(a,b]$; the same statement holds on $\mathbb{R}^n$. The strict inclusion $\mathcal{M}_\lambda \subsetneq \mathcal{P}((0,1])$ is the content of the next subsection.

4.2 Non-measurable sets: the Vitali construction

We exhibit a set $H \subseteq (0,1]$ outside $\mathscr{M}_\lambda$. The construction uses the Axiom of Choice and addition modulo 1, which wraps $(0,1]$ into a circle.

Definition 4.4: Addition modulo 1

For $x, y \in (0,1]$ define

$$x \oplus y \;=\; \begin{cases} x+y & \text{if } x+y \le 1, \\ x+y-1 & \text{if } x+y > 1. \end{cases}$$

For $A \subseteq (0,1]$ and $x \in (0,1]$, set $A + x = \{y \in (0,1] : y - x \in A\}$, with subtraction taken mod 1.

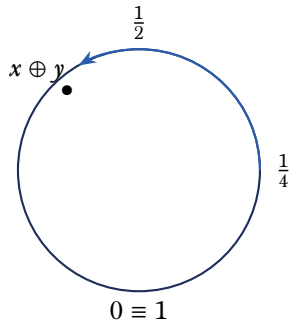


Figure 3. Addition mod 1 wraps $(0,1]$ onto a circle: shifts $A \mapsto A + x$ become rotations.

Lemma 4.5: Translation invariance on Borel sets

Let $\mathscr{L} = \{A \in \mathscr{M}_\lambda : \lambda(A + x) = \lambda(A) \text{ for all } x \in (0,1]\}$. Then $\mathscr{L}$ is a λ-system, and $\mathscr{A} \subseteq \mathscr{L}$ since $\lambda((a,b] + x) = \lambda((a + x, b + x]) = b - a$. By Dynkin's π-λ theorem, $\mathscr{B} = \sigma(\mathscr{A}) \subseteq \mathscr{L}$; equivalently, λ is translation invariant on every Borel set.

■ **Example 4.1 (Vitali set).** Define an equivalence relation on $(0,1]$ by $x \sim y \iff x - y \in \mathbb{Q}$; for instance $\frac{1}{\sqrt{2}} \sim \frac{1}{\sqrt{2}} + \frac{1}{100}$. By the Axiom of Choice pick a set $H \subseteq (0,1]$ containing exactly one representative from each equivalence class. Then for distinct $r_1, r_2 \in \mathbb{Q} \cap (0,1]$ we have $(H + r_1) \cap (H + r_2) = \varnothing$, and

$$(0,1] \;=\; \bigsqcup_{r \in \mathbb{Q} \cap (0,1]} (H + r).$$

If H were Lebesgue measurable, countable additivity together with Result 4.5 would give

$$1 \;=\; \lambda\big((0,1]\big) \;=\; \sum_{r \in \mathbb{Q} \cap (0,1]} \lambda(H + r) \;=\; \sum_{r \in \mathbb{Q} \cap (0,1]} \lambda(H),$$

which is 0 if $\lambda(H) = 0$ and ∞ if $\lambda(H) > 0$ — either way a contradiction. Hence $H \notin \mathscr{M}_\lambda$.

Remark 4.2. This shows the strict inclusion $\mathscr{M}_\lambda \subsetneq \mathscr{P}((0,1])$. A related fun fact: there is no infinite-dimensional analogue of Lebesgue measure: the only locally finite, translation-invariant Borel measure on an infinite-dimensional separable Banach space is the trivial one.

4.3 Product measures, briefly

The half-open construction generalises directly to $\mathbb{R}^p$ using half-open rectangles.

Definition 4.6: Lebesgue measure on $\mathbb{R}^p$

On $\mathbb{R}^p$, define

$$\lambda^{(p)}\big((a_1,b_1]\times\cdots\times(a_p,b_p]\big) \;=\; \prod_{i=1}^{p}\lambda\big((a_i,b_i]\big) \;=\; \prod_{i=1}^{p}(b_i-a_i).$$

The collection of half-open rectangles is a π-system, and the extension theorem gives a unique measure on $\mathcal{B}(\mathbb{R}^p)$. For $p=2$,

$$\lambda^{(2)}\big((a,b]\times(c,d]\big) \;=\; (b-a)(d-c) \;=\; \lambda\big((a,b]\big)\,\lambda\big((c,d]\big).$$

Definition 4.7: Product measure

Given two σ-finite measure spaces $(\mathbb{X},\mathcal{X},\mu)$ and $(\mathbb{Y},\mathcal{Y},\nu)$, the product measure space is $(\mathbb{X}\times\mathbb{Y},\,\mathcal{X}\times\mathcal{Y},\,\pi)$, where the product σ-field is

$$\mathcal{X}\times\mathcal{Y} \;=\; \sigma\big(\{A\times B : A\in\mathcal{X},\,B\in\mathcal{Y}\}\big),$$

and π is uniquely determined by

$$\pi(A\times B) \;=\; \mu(A)\,\nu(B), \qquad A\in\mathcal{X},\;B\in\mathcal{Y}.$$

Remark 4.3. The product of two Borel σ-fields satisfies $\mathcal{B}(\mathbb{X})\times\mathcal{B}(\mathbb{Y})\subseteq\mathcal{B}(\mathbb{X}\times\mathbb{Y})$, and equality holds in "nice" (e.g. second-countable) settings, including $\mathbb{X}=\mathbb{Y}=\mathbb{R}$.

4.4 Independence

Switch perspective from measure theory to probability: let $(\Omega,\mathcal{F},\mu)$ be a probability space.

Definition 4.8: Independence for sets

A countable collection $\{A_i\}_{i\in I}\subseteq\mathcal{F}$ is independent if, for every finite $J\subseteq I$,

$$\mu\left(\bigcap_{j\in J}A_j\right) \;=\; \prod_{j\in J}\mu(A_j).$$

■ **Example 4.2 (Standard 52-card deck).** Draw one card uniformly at random and let $A_1=\{\text{red}\}$, $A_2=\{\text{heart or club}\}$, $A_3=\{\text{queen}\}$. Then

$$\mu(A_1)=\tfrac{1}{2},\quad \mu(A_2)=\tfrac{1}{2},\quad \mu(A_3)=\tfrac{1}{13},$$

and one checks $\mu(A_1\cap A_2)=\tfrac{1}{4}$, $\mu(A_1\cap A_3)=\tfrac{1}{26}$, $\mu(A_2\cap A_3)=\tfrac{1}{26}$, $\mu(A_1\cap A_2\cap A_3)=\tfrac{1}{52}$, so the three events are independent.

Definition 4.9: Independence for σ-fields

A countable collection $\{\mathcal{F}_i\}_{i\in I}$ of sub-σ-fields of $\mathcal{F}$ is independent if every selection $\{A_i \in \mathcal{F}_i : i \in I\}$ is an independent collection of sets in the sense of Result 4.8.

The next theorem is the workhorse result: independence on a generating π-system already forces independence of the generated σ-fields.

Theorem 4.10: Independence from π-systems

Let $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{F}$ be π-systems. If

$$\mu(A_1 \cap A_2) \;=\; \mu(A_1)\,\mu(A_2) \quad \text{for all } A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2,$$

then $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent.

5 Lecture 5 — Simple and Measurable Functions

With Lebesgue measure in hand, we now begin populating the measure space with functions. The strategy is the one of Lecture 3 recycled: start with the simplest possible class of functions (finite-step indicators), define everything we want for them, and later extend by limits to a much wider class. Two parallel languages will run side by side throughout: a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ hosts simple random variables, while a generic measure space $(\Omega, \mathcal{F}, \mu)$ hosts simple functions. The structural content is identical.

5.1 Simple random variables

Definition 5.1: Simple random variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space ($\mathbb{P}(\Omega) = 1$). A simple random variable is a map $X : \Omega \to \mathbb{R}$ such that

- $X(\omega)$ takes only finitely many values $\{x_1, \dots, x_p\} \subseteq \mathbb{R}$;
- for each i, the level set $\{\omega \in \Omega : X(\omega) = x_i\}$ lies in $\mathcal{F}$.

Equivalently, given a finite $\mathcal{F}$-measurable partition $\{A_i\}_{i=1}^{p}$ of Ω (so $\bigsqcup_{i=1}^{p} A_i = \Omega$ and $A_i \cap A_j = \varnothing$ for $i \neq j$),

$$X(\omega) \;=\; \sum_{i=1}^{p} x_i \, \mathbf{1}[\omega \in A_i].$$

The probability that X hits a particular value reads off the partition directly:

$$\mathbb{P}(X = x_i) \;=\; \mathbb{P}\big(\{\omega \in \Omega : X(\omega) = x_i\}\big) \;=\; \mathbb{P}(A_i),$$

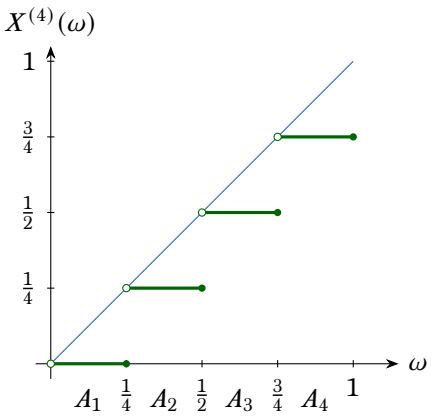
and the expectation is

$$\mathbb{E}X \;=\; \sum_{i=1}^{p} x_i \, \mathbb{P}(X = x_i).$$

■ **Example 5.1 (Staircase approximation of the identity on $(0,1]$).** Take $\Omega = (0,1]$ with Lebesgue measure, partition by $A_1 = (0, \tfrac14]$, $A_2 = (\tfrac14, \tfrac12]$, $A_3 = (\tfrac12, \tfrac34]$, $A_4 = (\tfrac34, 1]$, and set $x_i = (i-1)/4$. Each cell has $\lambda(A_i) = \tfrac14$, so the simple random variable

$$X^{(4)}(\omega) \;=\; \sum_{i=1}^{4} \tfrac{i-1}{4} \, \mathbf{1}[\omega \in A_i]$$

takes the values $0, \tfrac14, \tfrac12, \tfrac34$ each with probability $\tfrac14$, and $\mathbb{E}X^{(4)} = (0 + \tfrac14 + \tfrac12 + \tfrac34)/4 = 0.375$. Refining the partition into 2^m equal pieces produces a sequence of simple random variables $X^{(2^m)}$ that approximates the identity $\omega \mapsto \omega$ on $(0,1]$ ever more closely.



Remark 5.1. Letting the number of partition pieces grow to infinity, the simple random variables $X^{(2^m)}$ converge to the uniform distribution on $(0,1]$. The mode of convergence will be made precise in a later lecture.

5.2 Simple functions on a measure space

The same definition works verbatim with $\mathbb{P}$ replaced by a general measure μ; the only thing we lose is the probabilistic reading $\mathbb{P}(X = x_i)$.

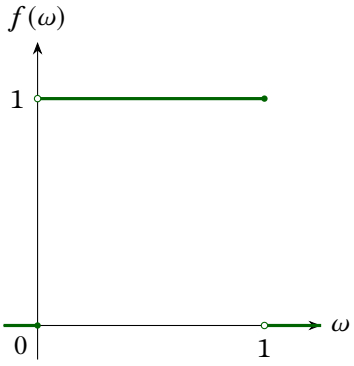
Definition 5.2: Simple function

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A function $f : \Omega \to \mathbb{R}$ is simple if there exist real numbers $x_1, \dots, x_p$ and sets $B_1, \dots, B_p \in \mathcal{F}$ such that

$$f(\omega) \;=\; \sum_{i=1}^{p} x_i \, \mathbf{1}[\omega \in B_i].$$

The sets B_i need not be disjoint, but a representation with disjoint B_i always exists by refining the family.

■ **Example 5.2 (Indicator of an interval).** For $\Omega = (0,1]$ with Borel σ-field and Lebesgue measure, the function $f(\omega) = \mathbf{1}[\omega \in (0,1]]$ is simple: it takes the value 1 on $(0,1]$ and 0 elsewhere.



Proposition 5.3: Algebra of simple functions

If $f, g : \Omega \to \mathbb{R}$ are simple functions, then so are

$$f + g, \qquad f \cdot g, \qquad \max\{f, g\}, \qquad \min\{f, g\}.$$

The integral of a simple function is what one would write down by hand: value × size of the level set, summed.

Definition 5.4: Integral of a simple function

For a simple function $f = \sum_{i=1}^{p} x_i \, \mathbf{1}[\,\cdot \in B_i\,]$ on $(\Omega, \mathcal{F}, \mu)$ with disjoint B_i, the integral of f with respect to μ is

$$\int f \, d\mu \;:=\; \sum_{i=1}^{p} x_i \, \mu(B_i).$$

Remark 5.2. For non-negative simple f, g and a scalar $c > 0$ the integral is linear: $\int (f + g)\, d\mu = \int f \, d\mu + \int g \, d\mu$ and $\int cf \, d\mu = c \int f \, d\mu$. This is the seed from which the Lebesgue integral on a much wider class of functions will grow in the next lectures.

5.3 Measurable functions

To go beyond finite-valued functions, we drop the requirement that f take only finitely many values and ask instead that the preimages of nice sets be measurable.

Definition 5.5: Measurable function

Let $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{Y}, \mathcal{Y})$ be measurable spaces. A function $f : \mathbb{X} \to \mathbb{Y}$ is $\mathcal{X}/\mathcal{Y}$-measurable if

$$f^{-1}(B) \;\in\; \mathcal{X} \qquad \text{for every } B \in \mathcal{Y},$$

where $f^{-1}(B) = \{x \in \mathbb{X} : f(x) \in B\}$.

Remark 5.3. When $(\mathbb{Y}, \mathcal{Y}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we call f Borel measurable. Replacing $\mathcal{B}(\mathbb{R})$ by the Lebesgue σ-field $\mathcal{M}_\lambda(\mathbb{R})$ gives the strictly larger class of Lebesgue measurable functions. Measurable random variables on a probability space are exactly measurable functions $X : \Omega \to \mathbb{R}$.

Remark 5.4. Inverse images preserve set operations:

$$f^{-1}\Big(\bigcup_i A_i\Big) \;=\; \bigcup_i f^{-1}(A_i), \qquad f^{-1}(\mathbb{Y} \setminus A) \;=\; \mathbb{X} \setminus f^{-1}(A).$$

A useful corollary: $\{f^{-1}(B) : B \in \mathcal{Y}\}$ is itself a σ-field on $\mathbb{X}$, and f is measurable iff this σ-field is contained in $\mathcal{X}$. To prove two σ-fields are equal, it suffices (as always) to verify both inclusions $A \subseteq B$ and $B \subseteq A$.

5.4 Useful facts about measurability

The following stability properties are the working toolkit. They say, roughly, that measurability survives every reasonable operation one might want to perform.

Proposition 5.6: Generators suffice

If $\mathcal{Y} = \sigma(\mathcal{A})$ for some collection $\mathcal{A}$, then f is $\mathcal{X}/\mathcal{Y}$-measurable iff $f^{-1}(A) \in \mathcal{X}$ for every $A \in \mathcal{A}$. In particular, since $\mathcal{B}(\mathbb{R})$ is generated by the half-lines $A_t = (-\infty, t]$ for $t \in \mathbb{R}$, a function $f : \mathbb{X} \to \mathbb{R}$ is Borel measurable iff

$$\{x \in \mathbb{X} : f(x) \le t\} \in \mathcal{X} \qquad \text{for every } t \in \mathbb{R}.$$

Proposition 5.7: Indicator functions

For any $A \in \mathcal{X}$, the indicator $f(x) = \mathbf{1}[x \in A]$ is measurable. The σ-field generated by f^{-1} is simply $\{\varnothing, A, A^c, \mathbb{X}\} \subseteq \mathcal{X}$.

Proposition 5.8: Algebra of measurable functions

For measurable $f, g : \mathbb{X} \to \mathbb{R}$, the functions $f + g$ and fg are measurable.

Proposition 5.9: Pointwise limits of measurable functions

For a sequence $\{f_i\}_{i=1}^{\infty}$ of measurable functions from $\mathbb{X}$ to $\mathbb{R}$, each of

$$\sup_i f_i, \quad \inf_i f_i, \quad \limsup_i f_i, \quad \liminf_i f_i$$

is measurable, and $\lim_i f_i$ is measurable wherever it exists. The key identity is

$$\{x : \sup_i f_i(x) \le t\} = \bigcap_i \{x : f_i(x) \le t\},$$

a countable intersection of measurable sets.

Proposition 5.10: Continuous functions are Borel measurable

Every continuous $f : \mathbb{R} \to \mathbb{R}$ (or, more generally, between topological spaces equipped with their Borel σ-fields) is Borel measurable. The reason: preimages of open sets under continuous maps are open, and open sets generate the Borel σ-field.

Remark 5.5. Given any collection $\{f_i\}_{i \in I}$ of functions $f_i : \mathbb{X} \to \mathbb{Y}$, one can always equip $\mathbb{X}$ with the smallest σ-field that makes every f_i measurable, namely $\sigma(\{f_i^{-1}(B) : i \in I, B \in \mathcal{Y}\})$. This is the canonical way to manufacture measurability rather than verify it.

5.5 Almost everywhere

A function-level analogue of "measure-zero exception" lets us identify functions that disagree only on a negligible set.

Definition 5.11: Almost everywhere / almost surely

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g : \Omega \to \mathbb{R}$. We say $f = g$ almost everywhere (written $f = g$ a.e.) if

$$\mu\big(\{\omega \in \Omega : f(\omega) \neq g(\omega)\}\big) = 0.$$

On a probability space the same notion is called almost surely (a.s.), or equivalently with probability one (wp1).

■ **Example 5.3 (Dirichlet-style equality).** On $\big((0,1], \mathcal{B}((0,1]), \lambda\big)$, set $f(t) = 0$ for all $t \in (0,1]$, and

$$g(t) = \begin{cases} 0 & t \in (0,1] \setminus \mathbb{Q}, \\ 1 & t \in (0,1] \cap \mathbb{Q}. \end{cases}$$

Then $f = g$ almost everywhere: the disagreement set is $(0,1] \cap \mathbb{Q}$, which is countable. Enumerate it as $\{q_1, q_2, \dots\}$ and cover q_m by the half-open interval $(q_m - 2^{-m}, q_m + 2^{-m+1}]$; the union of these intervals has Lebesgue measure at most $\sum_m 3 \cdot 2^{-m} < \infty$, and a sharper argument gives $\lambda\big((0,1] \cap \mathbb{Q}\big) = 0$.

6 Lecture 6 — Integration and Convergence Theorems

Lecture 5 introduced simple functions and defined their integral $\int s\,d\mu = \sum_i x_i \mu(B_i)$. We now extend the integral to arbitrary measurable $f : \Omega \to [-\infty, \infty]$ by approximation from below by simple functions, then state the three convergence theorems — Monotone Convergence, Fatou's Lemma, Dominated Convergence — that justify exchanging limits and integrals.

6.1 Lebesgue integration for measurable functions

We work on a measure space $(\Omega, \mathcal{F}, \mu)$ and consider measurable functions taking values in the extended real line $[-\infty, \infty]$. Allowing $\pm\infty$ is convenient because then sets like $f^{-1}(\infty)$ are well defined and limits of measurable functions stay measurable.

Definition 6.1: Increasing convergence $f_i \uparrow f$

For a sequence of measurable functions $f_i : \Omega \to [-\infty, \infty]$ and a measurable f, we write $f_i \uparrow f$ to mean

$$f_i(\omega) \le f_{i+1}(\omega) \quad \text{for all } \omega \in \Omega, \qquad \text{and} \qquad f_i(\omega) \longrightarrow f(\omega) \quad \text{as } i \to \infty.$$

The decreasing analogue $f_i \downarrow f$ is defined symmetrically.

The next result is the workhorse of the construction: every measurable function is a monotone limit of simple ones, and the dyadic recipe $f_i = 2^{-i}\lfloor 2^i f \rfloor$ is concrete enough to use in proofs.

Theorem 6.2: Approximation by simple functions

Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{A}$ be a π-system generating $\mathcal{F}$. Suppose $\mathcal{V}$ is a linear space of measurable functions such that

1. $\mathbf{1}_\Omega \in \mathcal{V}$ and $\mathbf{1}_A \in \mathcal{V}$ for every $A \in \mathcal{A}$;
2. whenever $f_i \in \mathcal{V}$ and $f_i \uparrow f$, one has $f \in \mathcal{V}$.

Then $\mathcal{V}$ contains every measurable function. In particular, for any non-negative measurable f, the simple functions

$$f_i \;=\; 2^{-i}\lfloor 2^i f \rfloor$$

satisfy $f_i \uparrow f$; a general measurable f is then handled via the decomposition $f = f^+ - f^-$ below.

Definition 6.3: Positive and negative parts

For a measurable $f : \Omega \to [-\infty, \infty]$, define

$$f^+(\omega) \;=\; \begin{cases} f(\omega) & \text{if } f(\omega) \ge 0, \\ 0 & \text{otherwise,} \end{cases} \qquad f^-(\omega) \;=\; \begin{cases} -f(\omega) & \text{if } f(\omega) \le 0, \\ 0 & \text{otherwise.} \end{cases}$$

Both f^+ and f^- are non-negative and measurable, and $f = f^+ - f^-$, $|f| = f^+ + f^-$.

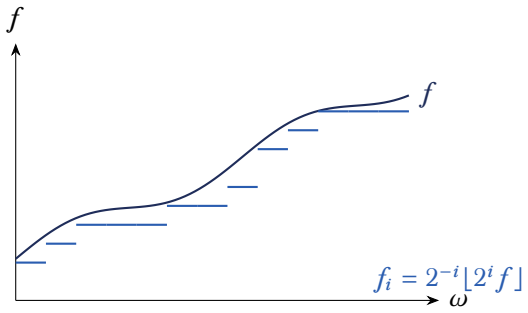


Figure 4. Dyadic simple approximation: rounding f down to the nearest multiple of 2^{-i} gives a simple function $f_i \le f$ with $f_i \uparrow f$ pointwise.

Definition 6.4: Integral of a measurable function

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.
Non-negative case. For a measurable $f : \Omega \to [0, \infty]$,

$$\int f \, d\mu \;=\; \sup\left\{ \int s \, d\mu \;:\; s \text{ simple}, \, 0 \le s \le f \right\} \;=\; \sup_{\text{partitions } \{A_i\} \text{ of } \Omega} \sum_i \left\{ \inf_{\omega \in A_i} f(\omega) \right\} \mu(A_i).$$

General case. For measurable $f : \Omega \to [-\infty, \infty]$,

$$\int f \, d\mu \;=\; \int f^+ \, d\mu \;-\; \int f^- \, d\mu,$$

provided not both terms are ∞. When both $\int f^+ d\mu$ and $\int f^- d\mu$ are finite we say f is integrable; the conventions $0 \cdot \infty = 0$ and $c \cdot \infty = \infty$ $(c > 0)$ keep the formula meaningful when f is supported on a set of infinite measure.

6.2 The three convergence theorems

The defining feature of the Lebesgue integral is that it interacts cleanly with limits. The next three theorems — each a statement about exchanging **lim** and $\int$ — are essentially the reason the construction is worth the trouble.

Theorem 6.5: Monotone convergence

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\{f_i\}_{i=1}^{\infty}$ be measurable functions with $f_i \uparrow f$ almost everywhere and $\int f_1 \, d\mu > -\infty$. Then

$$\int f_i \, d\mu \;\uparrow\; \int f \, d\mu.$$

Remark 6.1. The non-negativity hypothesis $f_i \ge 0$ commonly attached to the MCT is not needed once $\int f_1 \, d\mu > -\infty$: writing $h_i = f_i - f_1 \ge 0$ reduces the general case to the non-negative one. Convergence $f_i \uparrow f$ only needs to hold μ-almost everywhere (the bad set has measure zero and contributes nothing).

Theorem 6.6: Fatou's lemma

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\{f_i\}_{i=1}^{\infty}$ be non-negative measurable functions $\Omega \to \mathbb{R}$. Then

$$\int \liminf_{i\to\infty} f_i \, d\mu \;\le\; \liminf_{i\to\infty} \int f_i \, d\mu.$$

Remark 6.2. Setting $g_j = \inf_{i \ge j} f_i$ gives $g_j \uparrow \liminf_i f_i$ with $g_j \le f_i$ for $i \ge j$; MCT applied to (g_j) and monotonicity of the integral combine to give the inequality. The inequality is genuinely one-sided — the moving-bump example $f_i = \mathbf{1}_{[i,i+1]}$ on $(\mathbb{R}, \mathcal{B}, \lambda)$ has $\liminf f_i = 0$ yet $\int f_i \, d\mu = 1$ for every i.

Theorem 6.7: Dominated convergence

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, g a non-negative integrable function, and $\{f_i\}_{i=1}^{\infty}$ measurable with

$$|f_i(\omega)| \;\le\; g(\omega) \quad \text{for all } i \text{ and all } \omega \in \Omega, \qquad f_i(\omega) \;\longrightarrow\; f(\omega) \quad \text{for each } \omega \in \Omega.$$

Then f is integrable and

$$\int f_i \, d\mu \;\longrightarrow\; \int f \, d\mu.$$

Remark 6.3. The proof sandwiches f_i between the monotone envelopes $f_i^{\wedge} = \inf_{j \ge i} f_j \uparrow f$ and $f_i^{\vee} = \sup_{j \ge i} f_j \downarrow f$; MCT applied to $f_i^{\wedge} + g$ (increasing) and to $g - f_i^{\vee}$ (also increasing) gives

$$\int f_i^{\wedge} \, d\mu \;\le\; \int f_i \, d\mu \;\le\; \int f_i^{\vee} \, d\mu,$$

and both outer integrals converge to $\int f \, d\mu$.

Remark 6.4. The three theorems form a hierarchy: MCT is the foundation, Fatou is its immediate corollary via monotone envelopes from below, and DCT follows from Fatou applied to $g \pm f_i$. Each weakens the hypotheses of its predecessor (monotonicity → non-negativity → integrable domination) at the cost of requiring more setup.

7 Lecture 7 — Lebesgue–Stieltjes Measure; Fubini–Tonelli

We are halfway through the integration arc. Today's two themes both push measures between spaces. First: a measurable map $\psi\colon \mathbb{X} \to \mathbb{Y}$ carries a measure μ on $\mathbb{X}$ to its image measure $\nu = \mu \circ \psi^{-1}$ on $\mathbb{Y}$; when μ is Lebesgue measure on $\mathbb{R}$ and ψ is built from a distribution function F, this produces the Lebesgue–Stieltjes measure dF. Second: given two σ-finite spaces, one constructs the product measure on $\mathbb{X} \times \mathbb{Y}$, and the Fubini–Tonelli theorem licences swapping the order of integration.

7.1 Image measures and Lebesgue–Stieltjes

Definition 7.1: Image measure

Let $(\mathbb{X}, \mathcal{X}, \mu)$ and $(\mathbb{Y}, \mathcal{Y})$ be measurable spaces and $\psi\colon \mathbb{X} \to \mathbb{Y}$ measurable. The image (or push-forward) measure of μ under ψ is the set function $\nu = \mu \circ \psi^{-1}$ on $\mathcal{Y}$,

$$\nu(B) \;=\; \mu\big(\psi^{-1}(B)\big), \qquad B \in \mathcal{Y}.$$

Remark 7.1. Measurability of ψ is what makes $\psi^{-1}(B) \in \mathcal{X}$, so $\nu(B)$ is defined. Inverse images preserve countable unions and complements, so ν is automatically a measure on $\mathcal{Y}$. This is the construction that turns Lebesgue measure on $\mathbb{R}$ into the Lebesgue–Stieltjes measure attached to a distribution function.

Theorem 7.2: Lebesgue–Stieltjes measure

Let $F\colon \mathbb{R} \to \mathbb{R}$ be non-constant, right-continuous, and non-decreasing. There exists a unique measure dF on $\mathcal{B}(\mathbb{R})$ such that for all $a < b$ in $\mathbb{R}$,

$$dF\big((a,b]\big) \;=\; F(b) - F(a).$$

Writing $F(\infty) = \lim_{x\to\infty} F(x)$, $F(-\infty) = \lim_{x\to-\infty} F(x)$, $I = (F(-\infty), F(\infty))$, and

$$g(y) \;=\; \inf\{x \in \mathbb{R} : y \le F(x)\}, \qquad y \in I,$$

the measure dF is realised as the image of Lebesgue measure under g:

$$dF \;=\; \lambda \circ g^{-1}.$$

Lemma 7.3: Properties of the left-inverse g

With F and g as in Result 7.2, the function g is left-continuous and non-decreasing on I, and

$$g(y) \le x \iff y \le F(x) \qquad \text{for } y \in I,\ x \in \mathbb{R}.$$

Concretely, for fixed $y \in I$ the set $J_y = \{x \in \mathbb{R} : y \le F(x)\}$ equals $[g(y), \infty)$, so g is Borel measurable.

Remark 7.2. Once g is Borel measurable and left-continuous, $\lambda \circ g^{-1}$ is a measure on $\mathcal{B}(\mathbb{R})$ and

$$dF\big((a,b]\big) = \lambda\big(\{y : g(y) > a,\ g(y) \le b\}\big) = \lambda\big((F(a), F(b)]\big) = F(b) - F(a).$$

Uniqueness follows from the π-λ argument used for Lebesgue measure: any other measure μ

with $\mu((a,b]) = F(b) - F(a)$ must agree with dF on the π-system of half-open intervals, hence on every Borel set.

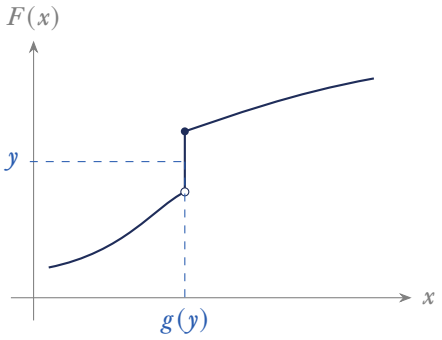


Figure 5. A right-continuous non-decreasing F with a jump. $g(y) = \inf\{x : y \le F(x)\}$ reads the picture sideways: a level y inside the jump still resolves to a single x-value.

Definition 7.4: Radon measure

A measure μ on $(\Omega, \mathcal{B})$, with $\mathcal{B}$ the Borel σ-field, is a Radon measure if $\mu(K) < \infty$ for every compact $K \in \mathcal{B}$.

Remark 7.3. dF is a Radon measure on $\mathbb{R}$, and conversely every non-zero Radon measure on $\mathcal{B}(\mathbb{R})$ can be written as $dF = \lambda \circ g^{-1}$ for some non-decreasing right-continuous F: take

$$F(x) \;=\; \begin{cases} \mu\big((0,x]\big) & \text{if } x \ge 0, \\ -\mu\big((x,0]\big) & \text{if } x < 0. \end{cases}$$

Then $F(b) - F(a) = \mu((a,b])$ for $a < b$, so $\mu = dF$ by uniqueness. Most measures one meets in practice are Radon.

7.2 Product σ-fields and the product measure

Switch settings: two measurable spaces, with the goal of building a measure on the Cartesian product.

Definition 7.5: Rectangles and product σ-field

Given measurable spaces $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{Y}, \mathcal{Y})$, a rectangle is a set $A \times B$ with $A \in \mathcal{X}$, $B \in \mathcal{Y}$. Write $\mathcal{R}$ for the collection of all rectangles. The product σ-field is

$$\mathcal{X} \times \mathcal{Y} \;=\; \sigma(\mathcal{R}).$$

Theorem 7.6: Existence and uniqueness of the product measure

Let $(\mathbb{X}, \mathcal{X}, \mu)$ and $(\mathbb{Y}, \mathcal{Y}, \nu)$ be σ-finite measure spaces, and let π be the set function on rectangles defined by

$$\pi(A \times B) \;=\; \mu(A)\,\nu(B), \qquad A \in \mathcal{X},\ B \in \mathcal{Y}.$$

Then π extends uniquely to a measure on $(\mathbb{X}\times\mathbb{Y}, \mathcal{X}\times\mathcal{Y})$, and for every $E\in\mathcal{X}\times\mathcal{Y}$,

$$\pi(E) \;=\; \iint \mathbf{1}_E(x,y)\,d\mu(x)\,d\nu(y) \;=\; \iint \mathbf{1}_E(x,y)\,d\nu(y)\,d\mu(x).$$

Remark 7.4. σ-finiteness cannot be dropped: without it, the extension above need not be unique. The strategy of proof is (i) prove the result for finite measures, then (ii) stitch together σ-finite exhaustions $\{A_i\}\subseteq\mathcal{X}$, $\{B_j\}\subseteq\mathcal{Y}$ with $\mu(A_i),\nu(B_j)<\infty$. The technical engine for step (i) is not Dynkin's π-λ theorem but its sibling, the monotone class theorem.

7.3 The monotone class theorem

The monotone class theorem is the natural "π-λ" tool when one starts from a field rather than a π-system.

Definition 7.7: Monotone class

A collection $\mathcal{M}$ of subsets of Ω is a monotone class if it is closed under monotone limits:

1. for $\{A_i\}_{i=1}^{\infty}\subseteq\mathcal{M}$ with $A_i\uparrow A=\bigcup_{i=1}^{\infty}A_i$, one has $A\in\mathcal{M}$;
2. for $\{A_i\}_{i=1}^{\infty}\subseteq\mathcal{M}$ with $A_i\downarrow A=\bigcap_{i=1}^{\infty}A_i$, one has $A\in\mathcal{M}$.

Recall that a field (algebra) on Ω is a non-empty collection containing Ω, closed under complements and finite unions. A field that is also a monotone class is automatically a σ-field.

Theorem 7.8: Monotone class theorem

Let $\mathcal{A}$ be a field on Ω and let $\mathcal{M}$ be a monotone class with $\mathcal{A}\subseteq\mathcal{M}$. Then

$$\sigma(\mathcal{A}) \;\subseteq\; \mathcal{M}.$$

Remark 7.5. Equivalently, the smallest monotone class containing a field $\mathcal{A}$ coincides with $\sigma(\mathcal{A})$. One uses this in the product-measure proof as follows: the rectangles $\mathcal{R}$ are a semi-ring, finite disjoint unions of rectangles form a field $\mathcal{A}$, and π is countably additive on $\mathcal{A}$. The monotone class theorem is what lifts "equality on $\mathcal{A}$" to "equality on $\sigma(\mathcal{A})=\mathcal{X}\times\mathcal{Y}$" for any two candidate extensions.

Lemma 7.9: Order of integration for indicators

Let $(\mathbb{X},\mathcal{X},\mu)$ and $(\mathbb{Y},\mathcal{Y},\nu)$ be finite measure spaces. Then for every $E\in\mathcal{X}\times\mathcal{Y}$,

$$\iint \mathbf{1}_E(x,y)\,d\mu(x)\,d\nu(y) \;=\; \iint \mathbf{1}_E(x,y)\,d\nu(y)\,d\mu(x).$$

The collection of E for which the displayed equality holds contains every rectangle $A\times B$ (since both sides equal $\mu(A)\nu(B)$), is closed under finite disjoint unions, and is a monotone class by Monotone Convergence. By Result 7.8 it equals $\mathcal{X}\times\mathcal{Y}$.

7.4 The Fubini–Tonelli theorem

The set-function identity in Result 7.6 extends to integrals of measurable functions, not just indicators.

Theorem 7.10: Fubini–Tonelli

Let $(\mathbb{X},\mathcal{X},\mu)$ and $(\mathbb{Y},\mathcal{Y},\nu)$ be σ-finite measure spaces, and let $f\colon \mathbb{X}\times\mathbb{Y}\to\mathbb{R}$ be $\mathcal{X}\times\mathcal{Y}$-measurable. Suppose either

$$f \geq 0 \qquad \text{(Tonelli)}, \qquad \text{or} \qquad \iint |f|\, d(\mu\times\nu) < \infty \quad \text{(Fubini)}.$$

Then

$$\int f\, d(\mu\times\nu) \;=\; \iint f(x,y)\, d\mu(x)\, d\nu(y) \;=\; \iint f(x,y)\, d\nu(y)\, d\mu(x).$$

Moreover $\int f(x,y)\, d\mu(x)$ is $\mathcal{Y}$-measurable in y, and $\int f(x,y)\, d\nu(y)$ is $\mathcal{X}$-measurable in x.

Remark 7.6. The proof slots together the three convergence theorems of Lecture 6 with the indicator case: the identity is immediate for indicators (Result 7.9), extends by linearity to simple functions, extends by Monotone Convergence to non-negative measurable f (Tonelli), and finally extends to integrable $f = f^+ - f^-$ by splitting positive and negative parts (Fubini). Both halves are needed in practice: Tonelli to check integrability by computing one of the iterated integrals of $|f|$; Fubini to then swap the order in the actual integral.

Remark 7.7. By induction the theorem extends to any finite product $\mathcal{X}_1\times\cdots\times\mathcal{X}_n$ of σ-finite spaces. The infinite product story is genuinely different: a countably infinite product of σ-finite spaces need not be σ-finite, but a countable product of probability spaces is again a probability space. We will return to this when constructing stochastic processes.

8 Lecture 8 — L^p Spaces and Classical Inequalities

Fix a measure space $(\Omega, \mathcal{F}, \mu)$. This lecture introduces the spaces $L^p(\Omega, \mathcal{F}, \mu)$, the four workhorse tail-bound inequalities (Markov, Chebyshev, Chernoff, Jensen) and the two inequalities (Hölder and Minkowski) that give L^p the structure of a normed vector space. We close with a density theorem that lets one approximate any L^p function by simple functions supported on sets of finite measure.

8.1 The spaces L^p

Definition 8.1: L^p space, $1 \le p < \infty$

For $p \in [1, \infty)$ and a measurable function $f : \Omega \to \mathbb{R}$, define

$$\|f\|_p \;=\; \left(\int_\Omega |f|^p \, d\mu \right)^{1/p}.$$

The space $L^p(\Omega, \mathcal{F}, \mu)$ consists of all measurable f with $\|f\|_p < \infty$, modulo equality μ-almost everywhere.

Definition 8.2: Essential supremum and L^∞

The essential supremum of a measurable function f is

$$\|f\|_\infty \;=\; \operatorname{ess\,sup} |f| \;=\; \inf\bigl\{ t \in [-\infty, \infty] : \mu(\{|f| > t\}) = 0 \bigr\}.$$

The space $L^\infty(\Omega, \mathcal{F}, \mu)$ consists of all measurable f with $\|f\|_\infty < \infty$, again modulo μ-a.e. equality.

Remark 8.1. The endpoint cases $p = 1$ and $p = \infty$ are honest analogues of the finite-p definition: as $p \to \infty$ one has $\|f\|_p \to \|f\|_\infty$ when μ is finite and f is bounded.

8.2 Markov, Chebyshev, Chernoff

The next three results all spring from the same one-line trick: bound the integrand below on the set $\{f \ge t\}$.

Theorem 8.3: Markov's inequality

Let $f \ge 0$ be measurable and $t > 0$. Then

$$\mu(\{f \ge t\}) \;\le\; \frac{1}{t} \int_\Omega f \, d\mu.$$

Corollary 8.4: Chebyshev's inequality

For any measurable f and $m \in \mathbb{R}$, $t > 0$,

$$\mu(\{|f - m| \geq t\}) \;\leq\; t^{-2} \int_{\Omega} (f - m)^2 \, d\mu.$$

Corollary 8.5: Chernoff's inequality

For any measurable f, $t \in \mathbb{R}$, and $\eta \geq 0$,

$$\mu(\{f \geq t\}) \;\leq\; e^{-\eta t} \int_{\Omega} e^{\eta f} \, d\mu.$$

Remark 8.2. Both Chebyshev and Chernoff follow from Result 8.3 applied to a non-negative transform: $(f - m)^2$ for Chebyshev, $e^{\eta f}$ for Chernoff. Chernoff is sharpest when one optimises over $\eta \geq 0$.

8.3 Convexity and Jensen's inequality

Definition 8.6: Convex function

Let $I \subseteq \mathbb{R}$ be an interval. A function $\phi \colon I \to \mathbb{R}$ is convex if for all $x, y \in I$ and $t \in [0,1]$,

$$\phi\big(tx + (1-t)y\big) \;\leq\; t\,\phi(x) + (1-t)\,\phi(y).$$

Geometrically: the secant line lies above the graph.

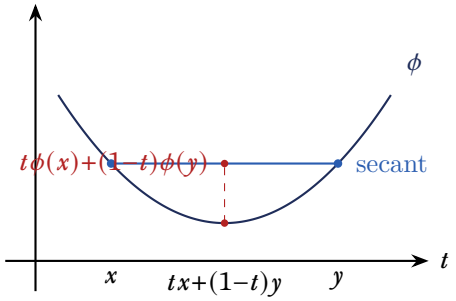


Figure 6. A convex function: the secant joining $(x, \phi(x))$ and $(y, \phi(y))$ lies above the graph.

Theorem 8.7: Jensen's inequality

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space (so $\mu(\Omega) = 1$), let $X \colon \Omega \to I$ be integrable with $\mathbb{E}X = \int X \, d\mu$ lying in the interior of I, and let $\phi \colon I \to \mathbb{R}$ be convex with $\mathbb{E}[\phi(X)]$ well defined. Then

$$\phi\big(\mathbb{E}X\big) \;\leq\; \mathbb{E}[\phi(X)].$$

Equivalently, for any measurable X and convex ϕ, $\phi\big(\int X \, d\mu\big) \leq \int \phi(X) \, d\mu$.

Remark 8.3. Convexity guarantees a supporting line: at $m = \mathbb{E}X$ one can choose $a, b \in \mathbb{R}$ with $\phi(x) \geq ax + b$ for all $x \in I$ and $\phi(m) = am + b$. Taking expectations of the inequality gives the

result; integrability of $\phi(X)$ follows because $\phi^-(x) \le |a|\,|x| + |b|$.

8.4 Hölder and Minkowski

We now turn to the two inequalities that pin down the geometry of L^p. Throughout, $p, q \in [1,\infty]$ are called conjugate exponents when

$$\frac{1}{p} + \frac{1}{q} \;=\; 1,$$

with the conventions $1/\infty = 0$ and $(p,q) \in \{(1,\infty),(\infty,1)\}$ included.

Theorem 8.8: Hölder's inequality

Let $p, q \in [1,\infty]$ be conjugate exponents and let f, g be measurable. Then

$$\|fg\|_1 \;=\; \int_\Omega |fg|\,d\mu \;\le\; \|f\|_p\,\|g\|_q.$$

In particular, if $f \in L^p$ and $g \in L^q$, then $fg \in L^1$.

Corollary 8.9: Cauchy–Schwarz

The choice $p = q = 2$ in Result 8.8 gives

$$\int_\Omega |fg|\,d\mu \;\le\; \left(\int_\Omega f^2\,d\mu\right)^{1/2}\left(\int_\Omega g^2\,d\mu\right)^{1/2} \;=\; \|f\|_2\,\|g\|_2.$$

Theorem 8.10: Minkowski's inequality

Let $p \in [1,\infty]$ and let f, g be measurable. Then

$$\|f+g\|_p \;\le\; \|f\|_p + \|g\|_p.$$

In particular, $L^p(\Omega,\mathcal{F},\mu)$ is closed under addition and $\|\cdot\|_p$ is a seminorm; on the quotient by μ-a.e. equality it is a norm.

Remark 8.4. For $p > 1$, Minkowski's bound is obtained by writing $|f+g|^p \le 2^{p-1}(|f|^p + |g|^p)$ (so $f+g \in L^p$) and then applying Hölder to the splitting $\int |f+g|^p \;=\; \int |f|\,|f+g|^{p-1} + \int |g|\,|f+g|^{p-1}$ with conjugate exponents p and $q = p/(p-1)$.

8.5 Approximation in L^p

The next theorem says that "simple functions supported on sets of finite measure" are dense in L^p for $p \in [1,\infty)$. It is the standard tool for reducing analytic statements to a check on indicator functions.

Theorem 8.11: Density of simple functions in L^p

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and assume there exist $A_n \in \mathcal{F}$ with $A_n \uparrow \Omega$ and $\mu(A_n) < \infty$ for all n (i.e. μ is σ-finite). Let

$$V_0 \;=\; \operatorname{span}\bigl\{\mathbf{1}_A : A \in \mathcal{F},\ \mu(A) < \infty\bigr\}$$

denote the simple functions supported on sets of finite measure. Then for every $p \in [1, \infty)$, $V_0 \subseteq L^p$ and for every $f \in L^p$ and every $\varepsilon > 0$ there exists $v \in V_0$ with

$$\|f - v\|_p \;<\; \varepsilon.$$

Remark 8.5. The proof has the usual three-step shape: (i) a Dynkin π-λ argument shows the class $\mathcal{L} = \{A \in \mathcal{F} : \mathbf{1}_A \text{ is approximable}\}$ is a λ-system containing the generating π-system, hence all of $\mathcal{F}$; (ii) for non-negative $f \in L^p$ the truncations $f_n = \min(n, 2^{-n}\lfloor 2^n f \rfloor)$ satisfy $|f - f_n|^p \to 0$ pointwise with $|f - f_n|^p \le |f|^p$, so dominated convergence gives $\|f - f_n\|_p \to 0$; (iii) general f is handled by splitting $f = f^+ - f^-$ and restricting to the exhausting sets A_n.

9 Lecture 9 — Convergence in Probability and Measure

We now study what it means for a sequence of probability measures, or of random variables, to converge. There are several inequivalent notions; this lecture introduces the four standard modes for random variables (almost sure, in probability, in L^p, in distribution), together with their parent notion at the level of measures (weak convergence). The Portmanteau theorem packages the equivalent characterisations of weak convergence, and a small Hasse diagram records the implications between the modes.

9.1 Weak convergence of probability measures

Let $(\Omega, \mathcal{F})$ be a measurable space and let $\{\mathbb{P}_i\}_{i=1}^{\infty}$ be a sequence of probability measures on $(\Omega, \mathcal{F})$. What should "$\mathbb{P}_i \to \mathbb{P}$" mean? The naive choice "$\mathbb{P}_i(A) \to \mathbb{P}(A)$ for every $A \in \mathcal{F}$" (setwise convergence) is too strong to be useful in practice; the standard notion fixes a topology on Ω and tests against continuous functions.

Definition 9.1: Weak convergence of measures

Let S be a metric space with Borel σ-field $\mathcal{S} = \mathcal{B}(S)$, and let $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$ be probability measures on $(S, \mathcal{S})$. We say $\mathbb{P}_i$ converges weakly to $\mathbb{P}$, written $\mathbb{P}_i \Rightarrow \mathbb{P}$, if

$$\int_S f \, d\mathbb{P}_i \longrightarrow \int_S f \, d\mathbb{P} \qquad \text{for every } f \in \mathcal{C}_B(S),$$

where $\mathcal{C}_B(S)$ denotes the bounded continuous real-valued functions on S.

Remark 9.1. Weak convergence is the topology of closeness between measures generated by the metric on S. Concretely, an ε-neighbourhood of $\mathbb{P}$ is determined by a finite collection $f_1, \dots, f_n \in \mathcal{C}_B(S)$: the neighbourhood is the set of all probability measures Q with $\left|\int f_i \, d\mathbb{P} - \int f_i \, dQ\right| < \varepsilon$ for each i. The weakest test against $f \in \mathcal{C}_B(S)$ gives the weakest of several useful metrics on the space of probability measures.

The next theorem lists the equivalent characterisations of weak convergence; it is the standard packaging that one cites whenever weak convergence appears in the wild.

Theorem 9.2: Portmanteau theorem

Let $\mathbb{P}$ and $\{\mathbb{P}_i\}_{i=1}^{\infty}$ be probability measures on a metric space $(S, \mathcal{S})$. The following are equivalent.

1. $\mathbb{P}_i \Rightarrow \mathbb{P}$, i.e. $\int f \, d\mathbb{P}_i \to \int f \, d\mathbb{P}$ for every $f \in \mathcal{C}_B(S)$.
2. $\int f \, d\mathbb{P}_i \to \int f \, d\mathbb{P}$ for every bounded uniformly continuous f.
3. $\limsup_i \mathbb{P}_i(C) \le \mathbb{P}(C)$ for every closed $C \subseteq S$.
4. $\liminf_i \mathbb{P}_i(U) \ge \mathbb{P}(U)$ for every open $U \subseteq S$.
5. $\lim_i \mathbb{P}_i(A) = \mathbb{P}(A)$ for every $A \in \mathcal{S}$ with $\mathbb{P}(\partial A) = 0$, where $\partial A = \bar{A} \cap \overline{A^c}$ is the topological boundary.

Remark 9.2. The implications (1) $\Rightarrow$ (2) and (2) $\Rightarrow$ (3) are the workhorse direction. For (2) $\Rightarrow$ (3) the trick is to take $C_\delta = \{x \in S : d(x, C) < \delta\}$ and choose a uniformly continuous f with $f = 1$ on C and $f = 0$ off C_δ (Urysohn's lemma); since $C_\delta \downarrow C$ as $\delta \to 0^+$, one has $\mathbb{P}_i(C) \le \int f \, d\mathbb{P}_i \to$

$\int f\,d\mathbb{P} \le \mathbb{P}(C_\delta) < \mathbb{P}(C) + \varepsilon$, and $\varepsilon \downarrow 0$ finishes the argument.

Remark 9.3. Changing the test class for f tightens the convergence:

- $f \in \mathscr{C}_B(S)$ is weak convergence;
- $\sup_f \left|\int f\,d\mathbb{P}_i - \int f\,d\mathbb{P}\right| \to 0$ over all continuous $f\colon S \to [-1,1]$ is the Radon metric;
- the same supremum over all measurable $f\colon S \to [-1,1]$ gives the total variation distance;
- the supremum restricted to 1-Lipschitz $f\colon S \to [-1,1]$ is the 1-Wasserstein distance, central to optimal transport.

9.2 Random variables and their distributions

We now lift the picture from measures to random variables. Fix a probability space $(\Omega,\mathcal{F},\mu)$ and a metric space $(S,\mathcal{S})$ (with $\mathcal{S} = \mathcal{B}(S)$). A random variable is a measurable map $X\colon \Omega \to S$; its distribution is the pushforward

$$\mathbb{P}(A) \;=\; \mu\big(X^{-1}(A)\big), \qquad A \in \mathcal{S}.$$

With this in place, expectations have the change-of-variables form

$$\mathbb{E}[X] \;=\; \int_\Omega X(\omega)\,d\mu(\omega) \;=\; \int_S x\,d\mathbb{P}(x).$$

For a sequence $\{X_i\}_{i=1}^\infty$ we write $\mathbb{P}_i$ for the distribution of X_i, and when no confusion arises we abuse notation and write $\mathbb{P}_i(A)$ for $\mathbb{P}(X_i \in A)$.

9.3 Modes of convergence

We collect the four standard modes; throughout, $X, X_1, X_2, \dots$ are random variables on a common probability space taking values in a metric space (S,d).

Definition 9.3: Convergence in distribution

X_i converges to X in distribution, written $X_i \xrightarrow{d} X$, if the laws $\mathbb{P}_i$ of X_i converge weakly to the law $\mathbb{P}$ of X: $\mathbb{P}_i \Rightarrow \mathbb{P}$ (Result 9.1).

Definition 9.4: Convergence in probability

X_i converges to X in probability, written $X_i \xrightarrow{\mathbb{P}} X$, if for every $\varepsilon > 0$,

$$\mu\big(\{\omega \in \Omega : d(X_i(\omega), X(\omega)) > \varepsilon\}\big) \;\longrightarrow\; 0 \qquad \text{as } i \to \infty.$$

In shorthand, $\mathbb{P}\big(d(X_i,X) > \varepsilon\big) \to 0$ for every $\varepsilon > 0$.

Definition 9.5: Almost sure convergence

X_i converges to X almost surely (or μ-almost everywhere), written $X_i \xrightarrow{\text{a.s.}} X$, if

$$\mu\big(\{\omega \in \Omega : X_i(\omega) \to X(\omega)\}\big) \;=\; 1,$$

or equivalently $\mu(\{\omega : X_i(\omega) \not\to X(\omega)\}) = 0$. This is pointwise convergence except on a μ-null set; the metric d does not enter the statement.

Definition 9.6: Convergence in L^p

For $p \in [1,\infty)$, X_i converges to X in L^p, written $X_i \xrightarrow{L^p} X$, if

$$\mathbb{E}\big[d(X_i,X)^p\big] \;=\; \int_\Omega d\big(X_i(\omega),X(\omega)\big)^p \, d\mu(\omega) \;\longrightarrow\; 0.$$

When $S = \mathbb{R}$ this reduces to $\int |X_i - X|^p \, d\mu \to 0$.

9.4 Hierarchy of convergence

The four modes are not equivalent; they are linked by a small lattice of implications.

Proposition 9.7: Hierarchy of modes

For random variables on a probability space $(\Omega,\mathcal{F},\mu)$:

1. $X_i \xrightarrow{\text{a.s.}} X \implies X_i \xrightarrow{\mathbb{P}} X$.
2. $X_i \xrightarrow{\mathbb{P}} X \implies X_i \xrightarrow{d} X$.
3. For any $p \in [1,\infty]$, $X_i \xrightarrow{L^p} X \implies X_i \xrightarrow{\mathbb{P}} X$.
4. For $1 \le q < p \le \infty$, $X_i \xrightarrow{L^p} X \implies X_i \xrightarrow{L^q} X$ (on a probability space; this uses Jensen).

None of the converses hold without extra hypotheses; in particular, a.s. convergence and L^p convergence are incomparable.

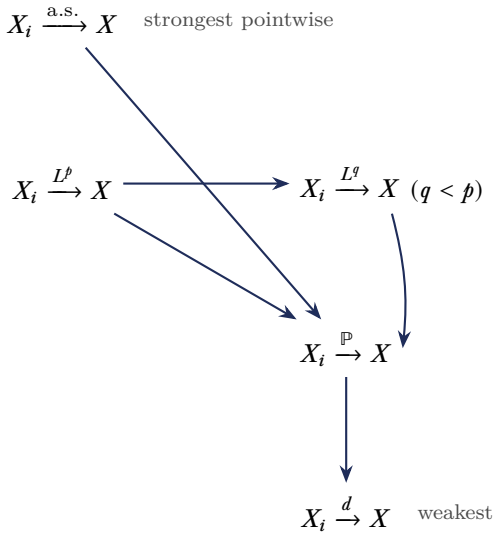


Figure 7. Hasse diagram of the four modes of convergence on a probability space. Arrows point from stronger to weaker; the horizontal arrow is the L^p-monotonicity from Jensen's inequality. There is no arrow between $\xrightarrow{\text{a.s.}}$ and $\xrightarrow{L^p}$ without extra integrability.

Remark 9.4. The implication $\xrightarrow{L^p} \Rightarrow \xrightarrow{\mathbb{P}}$ is a one-line consequence of Result 8.3: for any $\varepsilon > 0$,

$$\mu\big(d(X_i, X) > \varepsilon\big) \;\le\; \frac{\mathbb{E}[d(X_i, X)^p]}{\varepsilon^p} \;\longrightarrow\; 0.$$

The L^p-monotonicity uses Jensen applied to the convex map $t \mapsto t^{p/q}$ on a probability space: $\mathbb{E}|Y|^q \le (\mathbb{E}|Y|^p)^{q/p}$. The direction $\xrightarrow{\mathbb{P}} \Rightarrow \xrightarrow{d}$ is a corollary of the Portmanteau theorem (Result 9.2); the direction $\xrightarrow{\text{a.s.}} \Rightarrow \xrightarrow{\mathbb{P}}$ is dominated convergence applied to the indicator $\mathbf{1}_{\{d(X_i,X)>\varepsilon\}}$.

■ **Example 9.1 (Why a.s. and L^p are incomparable).** On $([0,1], \mathcal{B}, \lambda)$, set $X_n = n\,\mathbf{1}_{(0,1/n)}$. Then $X_n \to 0$ pointwise (so $X_n \xrightarrow{\text{a.s.}} 0$) but $\mathbb{E}[X_n] = 1$ for every n, so X_n does not converge to 0 in L^1. Conversely, the "typewriter" sequence of indicators of dyadic sub-intervals of $[0,1]$ satisfies $X_n \xrightarrow{L^p} 0$ for every p yet fails to converge at any single ω.

10 Lecture 10 — Hierarchy of Convergence; Borel–Cantelli; Prohorov

Lecture 9 set up the four modes of convergence for random variables and the equivalent formulations of weak convergence (Portmanteau). The job now is twofold: assemble these modes into a single hierarchy of implications, and develop the Borel–Cantelli lemmas — the standard tool for promoting summable bounds on $\mu(A_i)$ into almost-sure statements about whether A_i occurs only finitely often. We close with a brief look at Prohorov's theorem, which gives a compactness criterion for sequences of probability measures and underlies the classical proof of the central limit theorem.

10.1 Stronger metrics on the space of probability measures

Weak convergence $\mathbb{P}_i \Rightarrow \mathbb{P}$ (Result 9.1) is the weakest of a family of distance-like notions on probability measures, all of the form $\sup_{f\in\mathcal{F}} \left|\int f\, d\mathbb{P}_i - \int f\, d\mathbb{P}\right|$ for some test class $\mathcal{F}$. Enlarging the test class $\mathcal{F}$ yields a finer notion of closeness:

- $\mathcal{F} = C_b(S)$ (continuous bounded): weak convergence.
- $\mathcal{F} = \{f : S \to [-1,1] \text{ continuous}\}$: the Radon metric.
- $\mathcal{F} = \{f : S \to [-1,1] \text{ measurable}\}$: the total variation distance.
- $\mathcal{F} = \{f : S \to \mathbb{R} \text{ Lipschitz with constant } 1\}$: the 1-Wasserstein distance, central to optimal transport and machine learning, which quantifies how quickly the convergence occurs rather than only that it does.

Remark 10.1. All four are equivalent on a finite S, and all imply weak convergence on any metric space; the converse fails in general. We work almost exclusively with weak convergence below.

10.2 Hierarchy of modes of convergence

Recall from Lecture 9 the four modes for random variables $X_i, X : (\Omega, \mathcal{F}, \mu) \to (S, \rho)$: convergence in distribution $X_i \xrightarrow{d} X$, in probability $X_i \xrightarrow{\mathbb{P}} X$ (Result 9.4), almost surely $X_i \xrightarrow{a.s.} X$ (Result 9.5), and in L^p $X_i \xrightarrow{L^p} X$ (Result 9.6). The implications between them form a strict hierarchy.

Theorem 10.1: Hierarchy of convergence

Let X_i, X be random variables on $(\Omega, \mathcal{F}, \mu)$ with values in a metric space (S, ρ). Then

$$X_i \xrightarrow{a.s.} X \implies X_i \xrightarrow{\mathbb{P}} X \implies X_i \xrightarrow{d} X,$$
$$X_i \xrightarrow{L^p} X \implies X_i \xrightarrow{\mathbb{P}} X \quad \text{for every } p \in [1,\infty].$$

None of the reverse implications holds, and a.s. and L^p convergence are not comparable: neither implies the other without an additional uniform integrability or domination hypothesis.

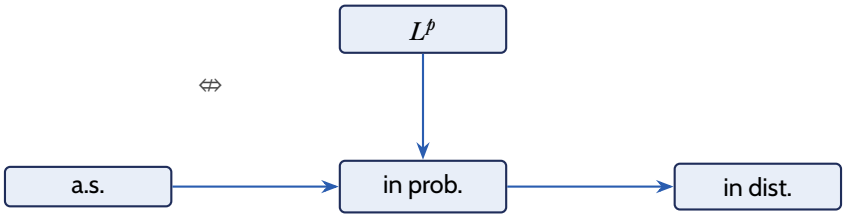


Figure 8. The four modes of convergence and the implications between them. Almost-sure and L^p convergence are not comparable.

10.3 Limsup, liminf, and the Borel–Cantelli setup

Fix a probability space $(\Omega, \mathcal{F}, \mu)$ and a sequence $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$. The set-theoretic limsup and liminf give a precise meaning to "A_i happens infinitely often" and "A_i happens eventually".

Definition 10.2: lim sup and lim inf of events

For $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$,

$$\limsup_i A_i = \bigcap_{i=1}^{\infty} \bigcup_{j \geq i} A_j, \qquad \liminf_i A_i = \bigcup_{i=1}^{\infty} \bigcap_{j \geq i} A_j.$$

We say A_i infinitely often (A_i i.o.) for $\limsup_i A_i$: $\omega \in \limsup_i A_i$ iff for every $N \in \mathbb{N}$ there exists $n \geq N$ with $\omega \in A_n$. We say A_i eventually (A_i ev.) for $\liminf_i A_i$: $\omega \in \liminf_i A_i$ iff there exists $N \in \mathbb{N}$ such that $\omega \in A_n$ for all $n \geq N$.

Remark 10.2. The two are dual in the sense $(\limsup_i A_i)^c = \liminf_i A_i^c$ and conversely, by De Morgan.

10.4 The Borel–Cantelli lemmas

The two lemmas are a one-sided pair: summability of $\mu(A_i)$ forces A_i to occur only finitely often almost surely; under the extra hypothesis of independence, divergence of the same series forces A_i to occur infinitely often almost surely.

Lemma 10.3: First Borel–Cantelli

Let $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$. If $\sum_{i=1}^{\infty} \mu(A_i) < \infty$, then

$$\mu\left(\limsup_i A_i\right) = 0.$$

Equivalently, with probability one only finitely many A_i occur. The proof is a one-line consequence of monotonicity and countable subadditivity: for every i,

$$\mu\left(\limsup_j A_j\right) \leq \mu\left(\bigcup_{j \geq i} A_j\right) \leq \sum_{j \geq i} \mu(A_j) \xrightarrow{i \to \infty} 0,$$

where the right-hand tail vanishes because the full series converges.

Lemma 10.4: Second Borel–Cantelli

Let $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ be independent. If $\sum_{i=1}^{\infty} \mu(A_i) = \infty$, then

$$\mu\left(\limsup_i A_i\right) = 1.$$

The argument is a complement-and-exponentiate trick. Independence of $\{A_i\}$ implies independence of $\{A_i^c\}$. For any $i \in \mathbb{N}$ and $k \geq i$,

$$\mu\left(\bigcap_{j=i}^{k} A_j^c\right) = \prod_{j=i}^{k}\left[1-\mu(A_j)\right] \leq \exp\left[-\sum_{j=i}^{k}\mu(A_j)\right],$$

using the elementary bound $1 - t \leq e^{-t}$ valid for all $t \in \mathbb{R}$. Letting $k \to \infty$ makes the right-hand side vanish, so $\mu(\bigcap_{j\geq i} A_j^c) = 0$ for every i; De Morgan then gives $\mu(\limsup_i A_i) = 1$.

Remark 10.3. Independence cannot be dropped from the second lemma: if $A_1 = A_2 = \cdots = A$ with $\mu(A) = \frac{1}{2}$, then $\sum \mu(A_i) = \infty$ but $\limsup_i A_i = A$ has probability $\frac{1}{2}$, not 1.

■ **Example 10.1 (Coin tosses produce every finite pattern).** Toss a fair coin independently and let A_i be the event that positions $i, i+1, \ldots, i+k-1$ spell out a fixed pattern of length k. Then $\mu(A_i) = 2^{-k}$, and the events $A_1, A_{k+1}, A_{2k+1}, \ldots$ are independent with $\sum_n \mu(A_{nk+1}) = \infty$. The second Borel–Cantelli lemma gives $\mu(A_{nk+1} \text{ i.o.}) = 1$: every finite pattern appears infinitely often almost surely.

10.5 Prohorov's theorem

The final piece of the convergence apparatus is a compactness criterion for sequences of probability measures: it is to weak convergence what Bolzano–Weierstrass is to bounded sequences in $\mathbb{R}^d$. The right notion of "boundedness" is tightness, capturing that no mass escapes to infinity.

Definition 10.5: Uniform tightness

A collection $\{\mu_i\}_{i\in I}$ of probability measures on a metric space (S, ρ) is uniformly tight if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subseteq S$ with

$$\mu_i(K_\varepsilon) > 1 - \varepsilon \qquad \text{for every } i \in I.$$

Theorem 10.6: Prohorov

Let $\{\mu_i\}_{i=1}^{\infty}$ be a sequence of probability measures on a metric space S. If $\{\mu_i\}$ is uniformly tight, then it is relatively sequentially compact for weak convergence: every subsequence μ_{i_k} admits a further subsequence $\mu_{i_{k_r}} \Rightarrow \mu$ for some probability measure μ (depending on the subsequence).

Remark 10.4. A useful subsubsequence corollary: if every subsequence of $\{\mu_i\}$ admits a further subsequence converging weakly to the same limit μ, then $\mu_i \Rightarrow \mu$. This is the classical route to the central limit theorem — one shows tightness, extracts a weak limit along a subsequence, and identifies the limit as the standard normal via characteristic functions.

11 Lecture 11 — Law of Large Numbers

The Borel–Cantelli machinery of Lecture 10 finally pays off. We fix a sequence $\{X_i\}_{i=1}^{\infty}$ of random variables on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, valued in $(\mathbb{R}, \mathcal{B})$, and ask in what sense the sample averages $n^{-1}S_n = n^{-1}\sum_{i=1}^{n} X_i$ approach the common mean. Two answers — one in probability under uncorrelation plus a second moment, one almost sure under independence and only a first moment — are the content of this lecture.

11.1 Setup: independence and identical distribution

Throughout, $X : \Omega \to \mathbb{R}$ is a random variable with law $\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$ for $A \in \mathcal{B}$, expectation $\mathbb{E}X = \int X(\omega)\, d\mathbb{P}$, and partial sums $S_n = \sum_{i=1}^{n} X_i$.

Definition 11.1: Independence of random variables

Random variables X and Y on $(\Omega, \mathcal{F}, \mathbb{P})$, valued in measurable spaces $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{Y}, \mathcal{Y})$ respectively, are independent if

$$\mathbb{P}\big(\{X \in A\} \cap \{Y \in B\}\big) \;=\; \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B) \qquad \text{for all } A \in \mathcal{X},\ B \in \mathcal{Y}.$$

The definition extends to a finite collection $\{X_i\}_{i=1}^{n}$ by requiring $\mathbb{P}\big(\bigcap_{i=1}^{n}\{X_i \in A_i\}\big) = \prod_{i=1}^{n}\mathbb{P}(X_i \in A_i)$. An infinite collection $\{X_i\}_{i=1}^{\infty}$ is independent if every finite subcollection is.

Remark 11.1. Since $\{X \in A\} = X^{-1}(A)$, independence of the random variables X and Y is equivalent to independence of the generated σ-fields $\sigma(X)$ and $\sigma(Y)$ in the sense of Result 4.9.

Definition 11.2: Identically distributed; i.i.d.

Random variables X and Y are identically distributed if the pushforward laws $\mathbb{P} \circ X^{-1}$ and $\mathbb{P} \circ Y^{-1}$ coincide on $\mathcal{B}$. A sequence $\{X_i\}_{i=1}^{\infty}$ is i.i.d. (independent and identically distributed) if it is independent and the X_i share a common law.

11.2 Weak law of large numbers

The weak law trades a strong moment hypothesis for a very mild dependence hypothesis: not full independence, only pairwise uncorrelation (a strictly weaker condition).

Theorem 11.3: Weak law of large numbers

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{X_i\}_{i=1}^{\infty}$ random variables with

$$\mathbb{E}X_i \;=\; c \in \mathbb{R}, \qquad \mathbb{E}X_i^2 \;=\; 1 \quad \text{for all } i,$$

and $\mathbb{E}\big[(X_i - c)(X_j - c)\big] = 0$ for all $i \neq j$. Then

$$\frac{S_n}{n} \;\xrightarrow{\;\mathbb{P}\;}\; c,$$

i.e. for every $\varepsilon > 0$, $\mathbb{P}\big(\big|n^{-1}S_n - c\big| \geq \varepsilon\big) \to 0$ as $n \to \infty$.

Remark 11.2. The proof reduces to $c = 0$ by replacing X_i with $X_i - c$, then applies Chebyshev: for any $t > 0$,

$$\mathbb{P}\left(\frac{|S_n|}{n} \geq t\right) \leq \frac{\mathbb{E}S_n^2}{t^2 n^2} = \frac{1}{t^2 n^2} \sum_{i,j=1}^{n} \mathbb{E}[X_i X_j] = \frac{1}{nt^2} \to 0,$$

where the cross terms vanish by uncorrelation and the diagonal sums to n by the unit second moment.

Remark 11.3. Uncorrelation is genuinely weaker than independence: independence of (X, Y) implies independence of $(f(X), g(Y))$ for any measurable f, g, hence $\mathrm{Cov}(f(X), g(Y)) = 0$ for every choice; uncorrelation asks this only for $f = g = \mathrm{id}$.

11.3 Strong law of large numbers

The strong law promotes "in probability" to "almost surely", removes the second moment hypothesis, but pays for it with full independence and identical distribution. Recall the variance

$$\mathrm{Var}(X) = \int (X - \mathbb{E}X)^2 \, d\mathbb{P}(\omega).$$

Theorem 11.4: Strong law of large numbers

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B})$. Then:

1. If $\mathbb{E}|X_1| < \infty$, then $\frac{S_n}{n} \xrightarrow{\text{a.s.}} c$ where $c = \mathbb{E}X_1$.
2. If $\mathbb{E}|X_1| = \infty$, then S_n/n does not converge to any finite limit (almost surely).

Remark 11.4. Compared with Result 11.3, no second-moment assumption is made on the X_i; only L^1 is needed. The trade-off is full independence (not just uncorrelation) and identical distribution. The "a.s." qualifier means convergence holds outside a $\mathbb{P}$-null set $N \subset \Omega$.

Remark 11.5. The divergence half (part 2) is the easier direction. The heuristic: if $\mathbb{E}|X_1| = \infty$ then $\sum_n \mathbb{P}(|X_n| > n) = \infty$, so by the second Borel–Cantelli lemma $|X_n| > n$ infinitely often. But on $\{n^{-1}S_n \to c\}$ one has $n^{-1}X_n = n^{-1}(S_n - S_{n-1}) \to 0$, contradicting $|X_n|/n > 1$ i.o.

Remark 11.6. The forward direction (part 1) is far more delicate. The standard route: reduce to $X_i \geq 0$ by writing $X_i = X_i^+ - X_i^-$ (independence of X, Y passes to X^+, Y^+), truncate $Y_i = X_i \mathbf{1}_{\{X_i \leq i\}}$ so that variances are finite, control the truncated partial sums $T_n = \sum_{i=1}^{n} Y_i$ along a geometric subsequence $k_n = \lfloor \delta^n \rfloor$ using Chebyshev plus the first Borel–Cantelli lemma, then sandwich the full sums S_i for $k_n \leq i \leq k_{n+1}$ and let $\delta \downarrow 1$.

■ **Example 11.1 (i.i.d. Bernoulli sample mean).** Let $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bernoulli}(p)$, so $\mathbb{E}X_i = p$ and $\mathrm{Var}(X_i) = p(1-p)$. Both moment hypotheses of the weak and strong laws are satisfied, so $n^{-1}S_n \to p$ both in probability (by Result 11.3) and almost surely (by Result 11.4). In particular, the empirical frequency of successes in n Bernoulli trials converges almost surely to the true success probability p — the formal statement behind the everyday claim "the average converges to the mean".

12 Lecture 12 — Central Limit Theorem; Characteristic Functions

The previous lecture closed the strong law of large numbers. We now turn to fluctuations: properly normalised, sums $S_n = X_1 + \cdots + X_n$ of iid mean-zero random vectors converge in distribution to a Gaussian. The proof rests on three tools — uniform tightness and Prohorov's theorem (Result 10.6, already established), the characteristic function and its uniqueness, and Lévy's continuity lemma — from which the central limit theorem drops out by a Taylor expansion.

12.1 Gaussian measures

Definition 12.1: Gaussian measure on $\mathbb{R}$

A Borel measure γ on $(\mathbb{R}, \mathcal{B})$ is Gaussian with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$ if

$$\gamma\big((a,b]\big) \;=\; \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\!\left[-\frac{1}{2\sigma^2}(x-m)^2\right] d\lambda(x).$$

For $\sigma = 0$ we set $\gamma = \delta_m$ (Dirac mass at m) and call γ a degenerate Gaussian measure.

Definition 12.2: Gaussian measure on $\mathbb{R}^d$

A Borel measure γ on $(\mathbb{R}^d, \mathcal{B})$ is Gaussian if for every linear functional $f\colon \mathbb{R}^d \to \mathbb{R}$ the induced measure $\gamma \circ f^{-1}$ on $(\mathbb{R}, \mathcal{B})$ is Gaussian. Equivalently, every linear combination of the coordinates is one-dimensional Gaussian.

Definition 12.3: Gaussian random variable

A random variable Z from a probability space $(\Omega, \mathcal{F}, \mu)$ to $(\mathbb{R}^d, \mathcal{B})$ is Gaussian if its law $\gamma := \mu \circ Z^{-1}$ is a Gaussian measure on $(\mathbb{R}^d, \mathcal{B})$.

Remark 12.1. For vectors $u, v \in \mathbb{R}^d$ we use the Euclidean inner product $\langle u, v\rangle = \sum_{i=1}^d u_i v_i$ and write $|u|^2 = \langle u, u\rangle$. A collection $\{X_i\}_{i=1}^{\infty}$ is iid if the X_i are pairwise independent and share a common law ("random variables induce measures").

12.2 Characteristic functions

The characteristic function is the Fourier transform of a probability measure; it linearises convolution and, by uniqueness below, encodes the measure completely.

Definition 12.4: Characteristic function

For a probability measure μ on $(\mathbb{R}^d, \mathcal{B})$, the characteristic function $\tilde{\mu}\colon \mathbb{R}^d \to \mathbb{C}$ is

$$\tilde{\mu}(t) \;:=\; \int \exp\!\big\{\, i\langle x, t\rangle \,\big\}\, d\mu(x).$$

When $\tilde{\mu}$ is integrable against Lebesgue measure on $\mathbb{R}^d$, the inverse transform recovers a density:

$$p(x) \;=\; (2\pi)^{-d} \int \tilde{\mu}(t)\, \exp\!\big\{\, -i\langle x, t\rangle \,\big\}\, d\lambda(t), \qquad \lambda\text{-a.e.},$$

with p the probability density function of μ.

Definition 12.5: Convolution of measures

For two measures μ, ν on $(\mathbb{R}^d, \mathcal{B})$, the convolution $\mu * \nu$ is the measure

$$(\mu * \nu)(B) \;:=\; \int \nu(B - x)\, d\mu(x), \qquad B \in \mathcal{B},$$

where $B - x = \{y \in \mathbb{R}^d : y + x \in B\}$. The operation $*$ is associative and commutative; the characteristic function of $\mu * \nu$ is $\tilde{\mu}\,\tilde{\nu}$; and if X, Y are independent with laws μ, ν, then $X + Y$ has law $\mu * \nu$.

Theorem 12.6: Uniqueness of characteristic functions

Let μ and ν be probability measures on $(\mathbb{R}^d, \mathcal{B})$. If $\tilde{\mu} = \tilde{\nu}$, then $\mu = \nu$.

Remark 12.2. The proof goes via Gaussian smoothing. Let γ_σ be the mean-zero Gaussian on $\mathbb{R}^d$ with covariance $\sigma^2 I$ and put $\mu^{(\sigma)} := \mu * \gamma_\sigma$, $\nu^{(\sigma)} := \nu * \gamma_\sigma$. The smoothed measures admit explicit densities

$$q^{(\sigma)}(x) \;=\; (2\pi)^{-d} \int \tilde{\nu}(t)\, \exp\!\big[-i\langle x, t\rangle - \tfrac{1}{2}\sigma^2 |t|^2\big]\, d\lambda(t),$$

and similarly for $p^{(\sigma)}$ with $\tilde{\mu}$. Hence $\tilde{\mu} = \tilde{\nu}$ forces $\mu^{(\sigma)} = \nu^{(\sigma)}$ for every $\sigma > 0$. Realising $\mu^{(\sigma)}$ as the law of $X + \sigma Z$ (with $X \sim \mu$, $Z \sim \gamma_1$ independent) and letting $\sigma \downarrow 0$ gives $X + \sigma Z \to X$ almost surely, hence in probability and so in distribution: $\mu^{(\sigma)} \Rightarrow \mu$, and likewise $\nu^{(\sigma)} \Rightarrow \nu$. Uniqueness of weak limits gives $\mu = \nu$.

12.3 Lévy's continuity lemma

Convergence of characteristic functions, plus tightness, controls weak convergence of the underlying measures.

Lemma 12.7: Lévy continuity

Let $\{\mu_i\}_{i=1}^{\infty}$ be a uniformly tight sequence of probability measures on $\mathbb{R}^d$. If the characteristic functions satisfy $\tilde{\mu}_i(v) \to \tilde{\mu}(v)$ for every $v \in \mathbb{R}^d$, then $\mu_i \Rightarrow \mu$, where μ is the (unique) probability measure with characteristic function $\tilde{\mu}$.

Remark 12.3. By Prohorov (Result 10.6), every subsequence μ_{i_k} has a further weakly convergent subsubsequence $\mu_{i_{k_r}} \Rightarrow \mu^*$. Continuity of the integrand forces $\widetilde{\mu^*} = \tilde{\mu}$ on all of $\mathbb{R}^d$, and uniqueness of characteristic functions (Result 12.6) identifies $\mu^* = \mu$. The standard subsubsequence trick (every subsequence has a further subsubsequence with the same weak limit) then promotes this to convergence of the full sequence.

12.4 The central limit theorem

We can now prove the headline result. The hypothesis is just iid plus a finite second moment.

Theorem 12.8: Central limit theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and let $\{X_n\}_{n=1}^{\infty}$ be iid random vectors on $(\mathbb{R}^d, \mathcal{B})$ with

$$\mathbb{E}X_n = 0 \qquad \text{and} \qquad \mathbb{E}\,|X_n|^2 < \infty.$$

Set $S_n = \sum_{j=1}^{n} X_j$. Then

$$n^{-\frac{1}{2}} S_n \xrightarrow{d} Z,$$

where Z is a Gaussian random vector on $\mathbb{R}^d$ with mean zero and covariance Σ given by $\Sigma_{jk} = \mathbb{E}[X_{nj} X_{nk}]$.

The strategy of the proof is a two-step: tightness of the normalised sums via a second-moment Chebyshev bound, and characteristic-function convergence via Taylor expansion. The two ingredients meet in Lévy's lemma.

Remark 12.4 (tightness via Chebyshev)**.** Since the X_j are mean zero and independent, $\mathbb{E}\langle X_j, X_k\rangle = 0$ for $j \neq k$, so

$$\mathbb{E}\left|n^{-\frac{1}{2}} S_n\right|^2 = \frac{1}{n}\,\mathbb{E}\left[\sum_{j,k=1}^{n} \langle X_j, X_k\rangle\right] = \mathbb{E}\,|X_j|^2.$$

For any $\varepsilon > 0$, choose $M_\varepsilon > 0$ with $\mathbb{E}|X_j|^2 / M_\varepsilon^2 < \varepsilon$; Chebyshev's inequality gives $\mathbb{P}(|n^{-\frac{1}{2}} S_n| > M_\varepsilon) < \varepsilon$, uniformly in n. The sequence $\{n^{-\frac{1}{2}} S_n\}$ is therefore uniformly tight.

Remark 12.5 (characteristic-function expansion)**.** Fix $v \in \mathbb{R}^d$. The scalars $\langle v, X_j\rangle$ are iid real-valued with $\mathbb{E}\langle v, X_j\rangle = 0$ and $\mathbb{E}\langle v, X_j\rangle^2 < \infty$. Define

$$h(v) := \mathbb{E}\exp\big(\, i\langle v, X_j\rangle\,\big).$$

Then $h(0) = 1$, $\nabla h(0) = 0$ and $\nabla^2 h(0) = -\Sigma$ where $\Sigma = \mathbb{E}[X_j X_j^{\top}]$. Taylor's theorem gives

$$h(v) = 1 - \tfrac{1}{2}\, v^{\top} \Sigma\, v + o(|v|^2).$$

Independence then yields, for any fixed v,

$$\mathbb{E}\exp\big\{\, i\langle n^{-\frac{1}{2}} S_n, v\rangle\,\big\} = h\big(n^{-\frac{1}{2}} v\big)^n = \left(1 - \frac{v^{\top}\Sigma v}{2n} + o\!\left(\frac{|v|^2}{n}\right)\right)^n \longrightarrow \exp\!\big\{-\tfrac{1}{2} v^{\top}\Sigma v\big\}$$

as $n \to \infty$. The right-hand side is the characteristic function of the mean-zero Gaussian Z on $\mathbb{R}^d$ with covariance Σ. Combining with tightness and Lévy's continuity (Result 12.7) gives $n^{-\frac{1}{2}} S_n \Rightarrow Z$ — convergence in distribution.

Remark 12.6. The covariance entry $\Sigma_{jk} = \mathbb{E}[X_{nj} X_{nk}]$ is independent of n by the iid hypothesis; the limiting Gaussian is the same regardless of which copy of X_n one uses to compute it. In the scalar case $d = 1$ the conclusion reduces to the familiar $n^{-1/2} S_n \Rightarrow \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \mathbb{E}X_1^2$.

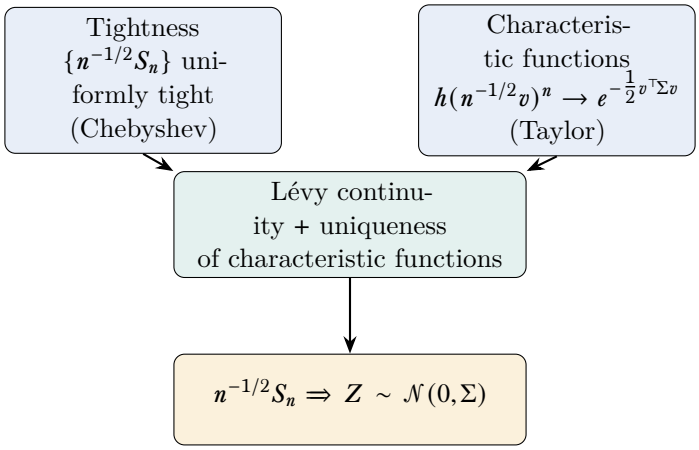


Figure 9. Architecture of the CLT proof: tightness and pointwise convergence of characteristic functions feed into Lévy's lemma; the limiting characteristic function identifies the Gaussian Z.

13 Lecture 13 — The Ergodic Theorem

The strong law of large numbers proved in Lecture 12 says that, for i.i.d. summands, time averages $n^{-1}S_n$ converge almost surely to the expected value. Ergodic theory generalises this picture to any measure-preserving dynamical system: replace "i.i.d." by "measure-preserving" and "$\mathbb{E}X_1$" by a conditional expectation on the σ-field of invariant sets. The two foundational results are Birkhoff's pointwise theorem (almost-sure convergence) and von Neumann's mean ergodic theorem (L^p convergence). Specialising to the shift on a product space recovers the SLLN.

13.1 Measure-preserving maps, invariance, ergodicity

Throughout this section $(\Omega, \mathcal{F}, \mu)$ is a measure space and $T : \Omega \to \Omega$ a measurable map. We are interested in time averages along the orbit $\omega, T\omega, T^2\omega, \dots$

Definition 13.1: Measure-preserving map

The map $T : \Omega \to \Omega$ is measure preserving if

$$\mu\big(T^{-1}(A)\big) \;=\; \mu(A), \qquad \text{for all } A \in \mathcal{F}.$$

Equivalently, the pushforward measure $\mu \circ T^{-1}$ coincides with μ: the dynamics does not distort the size of any measurable set.

Definition 13.2: Invariant set, invariant function

A set $A \in \mathcal{F}$ is T-invariant if $T^{-1}(A) = A$. The collection

$$\mathcal{F}_T \;=\; \{A \in \mathcal{F} : T^{-1}(A) = A\}$$

of all T-invariant sets is a σ-field. A measurable function $f : \Omega \to \mathbb{R}$ is invariant if $f = f \circ T$; equivalently, f is invariant if and only if it is $\mathcal{F}_T$-measurable.

Definition 13.3: Ergodic map

A measure-preserving map T is ergodic if every invariant set is trivial: for all $A \in \mathcal{F}_T$,

$$\mu(A) \;=\; 0 \quad \text{or} \quad \mu(A^c) \;=\; 0.$$

Equivalently, every T-invariant measurable function is constant μ-almost everywhere.

■ **Example 13.1 (Shift mod 1 on the circle).** On $\big((0,1], \mathcal{B}, \lambda\big)$ and a fixed $a \in (0,1]$, define the rotation

$$T(x) \;=\; x + a \mod 1 \;=\; \begin{cases} x + a & x + a \le 1, \\ x + a - 1 & x + a > 1. \end{cases}$$

T preserves Lebesgue measure: every half-open arc and its preimage have the same length. It is ergodic precisely when a is irrational.

■ **Example 13.2 (Baker's map).** On $(0,1]$ define $T(x) = 2x - \lfloor 2x \rfloor$. T is the doubling map; preimages of intervals split into two intervals of half the length, so Lebesgue measure is preserved. T is ergodic.

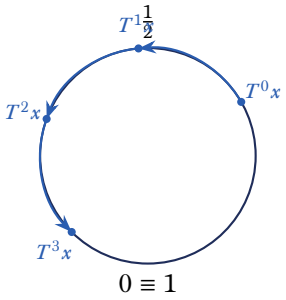


Figure 10. Orbit of a point under the rotation $T(x) = x + a \bmod 1$: for irrational a the orbit is dense, the dynamics is ergodic, and Birkhoff's theorem says time averages equal space averages.

The next two facts are the everyday tools used below; both follow directly from Results 13.1 and 13.3.

Proposition 13.4: Two basic facts

Let T be measure preserving on $(\Omega, \mathcal{F}, \mu)$.

1. If $f \in L^1(\Omega, \mathcal{F}, \mu)$ then $f \circ T \in L^1$ and
$$\int f \, d\mu = \int f \circ T \, d\mu.$$
2. If, in addition, T is ergodic and f is invariant, then $f = c$ μ-a.e. for some constant c.

13.2 Ergodic theorems

For the rest of the lecture, fix $(\Omega, \mathcal{F}, \mu)$ and a measure-preserving T. For $f : \Omega \to \mathbb{R}$ measurable set the Birkhoff sums

$$S_n = S_n(f) = f + f \circ T + f \circ T^2 + \cdots + f \circ T^{n-1}, \qquad S_0 \equiv 0.$$

Birkhoff's theorem controls the time averages $n^{-1} S_n(f)$ almost everywhere; von Neumann's controls them in L^p. Both rest on a single combinatorial estimate, the maximal ergodic lemma.

Lemma 13.5: Maximal ergodic lemma

Let $f \in L^1(\Omega, \mathcal{F}, \mu)$ and set $S^* = \sup_{n \geq 0} S_n(f)$. Then

$$\int_{\{S^* > 0\}} f \, d\mu \geq 0.$$

Theorem 13.6: Birkhoff's pointwise ergodic theorem

Let $(\Omega,\mathcal{F},\mu)$ be σ-finite, T measure preserving, and $f \in L^1(\Omega,\mathcal{F},\mu)$. There exists an invariant function $\bar f \in L^1(\Omega,\mathcal{F},\mu)$ with

$$\int |\bar f|\,d\mu \;\le\; \int |f|\,d\mu \qquad \text{and} \qquad \frac{S_n(f)}{n} \longrightarrow \bar f \quad \mu\text{-a.e. as } n\to\infty.$$

If T is ergodic and μ is a probability, then $\bar f = \int f\,d\mu$ almost everywhere.

Remark 13.1. The strategy is to show that $\liminf_n n^{-1}S_n(f)$ and $\limsup_n n^{-1}S_n(f)$ are both T-invariant and equal a.e. Invariance follows from

$$n^{-1}S_n(f)\circ T \;=\; n^{-1}\big[S_{n+1}(f)-f\big] \;=\; \frac{n+1}{n}\cdot\frac{S_{n+1}(f)}{n+1} \;-\; \frac{f}{n},$$

and one isolates the bad set

$$D_{a,b} \;=\; \Big\{\omega\in\Omega : \liminf_n n^{-1}S_n(f) < a < b < \limsup_n n^{-1}S_n(f)\Big\}$$

for rationals $a<b$. Each $D_{a,b}$ is T-invariant; an application of Result 13.5 to $g = f - b\,\mathbf{1}_B$ on a finite-measure subset $B\subseteq D_{a,b}$ yields

$$b\,\mu(D_{a,b}) \;\le\; \int_{D_{a,b}} f\,d\mu \;\le\; a\,\mu(D_{a,b}),$$

and $a<b$ forces $\mu(D_{a,b})=0$. Taking the countable union over rationals gives convergence in $[-\infty,\infty]$ on a full-measure set; the integrability bound $\int|\bar f|\,d\mu\le\int|f|\,d\mu$ falls out of Fatou's lemma applied to $n^{-1}|S_n(f)|$.

Theorem 13.7: von Neumann's mean ergodic theorem

Suppose $\mu(\Omega)<\infty$ and $p\in[1,\infty)$. For every $f\in L^p(\Omega,\mathcal{F},\mu)$ there exists $\bar f\in L^p$ such that

$$\frac{S_n(f)}{n} \longrightarrow \bar f \qquad \text{in } L^p.$$

Remark 13.2. The argument is a three-epsilon truncation. Because T is measure-preserving, $\|f\circ T^n\|_p = \|f\|_p$, so by Minkowski $\|n^{-1}S_n(f)\|_p\le\|f\|_p$. Given $\varepsilon>0$, choose $C>0$ and set $g=\min\{\max\{-C,f\},C\}$; then $\|f-g\|_p<\varepsilon/3$ and g is bounded by C, so dominated convergence upgrades the a.e. convergence $n^{-1}S_n(g)\to\bar g$ of Result 13.6 to L^p convergence. Fatou applied to $|n^{-1}S_n(f-g)|^p$ gives $\|\bar f-\bar g\|_p\le\|f-g\|_p$, and the triangle inequality

$$\Big\|\tfrac{S_n(f)}{n}-\bar f\Big\|_p \;\le\; \Big\|\tfrac{S_n(f-g)}{n}\Big\|_p + \Big\|\tfrac{S_n(g)}{n}-\bar g\Big\|_p + \|\bar g-\bar f\|_p \;<\; \varepsilon$$

finishes the proof.

13.3 Application: the strong law of large numbers, again

The two ergodic theorems give an almost free derivation of the SLLN by running the canonical i.i.d. construction through the shift map.

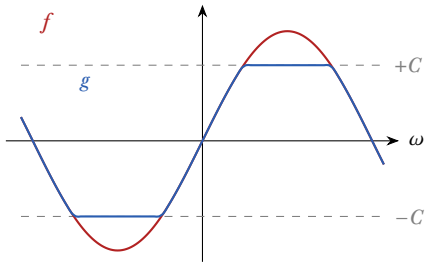


Figure 11. Truncation step in von Neumann's proof: the unbounded f (red) is clipped to a bounded $g = \min\{\max\{-C, f\}, C\}$ (blue); the tails are absorbed in $\|f - g\|_p < \varepsilon/3$, and dominated convergence handles g.

Let $(\Omega, \mathcal{F}, P)$ be a probability space carrying i.i.d. real-valued random variables $\{X_i\}_{i=1}^{\infty}$ with common distribution F. Set $(S, \mathcal{S}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{S})$ where $\mathcal{S}$ is generated by the π-system of cylinder sets

$$\mathcal{A} \;=\; \Big\{ \textstyle\prod_{n\in\mathbb{N}} A_n \;:\; A_n \in \mathcal{B}(\mathbb{R})\ \forall n,\ A_n = \mathbb{R} \text{ eventually} \Big\}.$$

The map $X : \Omega \to \mathbb{R}^{\mathbb{N}}$, $X(\omega) = (X_1(\omega), X_2(\omega), \dots)$, induces the product measure

$$\mu(A) \;=\; P \circ X^{-1}(A) \;=\; \prod_{n\in\mathbb{N}} dF(A_n), \qquad A = \textstyle\prod A_n.$$

Definition 13.8: Shift map on $\mathbb{R}^{\mathbb{N}}$

The shift map $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ drops the first coordinate:

$$T(x_1, x_2, x_3, \dots) \;=\; (x_2, x_3, x_4, \dots).$$

Proposition 13.9: The shift is measure-preserving and ergodic

Under the i.i.d. product measure μ above, the shift map T is measure preserving and ergodic. Ergodicity follows from Kolmogorov's zero-one law: every shift-invariant cylinder event lies in the tail σ-field $\bigcap_n \sigma(X_n, X_{n+1}, \dots)$ and so has probability 0 or 1.

Theorem 13.10: Strong law of large numbers, again

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. real-valued random variables with $\mathbb{E}|X_i| < \infty$. Then

$$\frac{S_n}{n} \;=\; \frac{X_1 + \cdots + X_n}{n} \;\xrightarrow{\text{a.s.}}\; \mathbb{E}X_i.$$

Remark 13.3. Take $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ to be the first-coordinate projection $f(x_1, x_2, \dots) = x_1$. With T the shift, $f \circ T^k(x) = x_{k+1}$, so the Birkhoff sums recover the partial sums:

$$S_n(f) \;=\; f + f \circ T + \cdots + f \circ T^{n-1} \;=\; X_1 + \cdots + X_n.$$

Result 13.6 gives an invariant $\bar{f} \in L^1$ with $n^{-1}S_n \to \bar{f}$ a.s. Since the shift is ergodic (Result 13.9), Result 13.3 forces $\bar{f}$ to be constant a.e.; identifying that constant via Result 13.7 at $p = 1$,

$$\bar{f} \;=\; \int \bar{f}\, d\mu \;=\; \lim_{n\to\infty} \int n^{-1} S_n(f)\, d\mu \;=\; \mathbb{E}X_i,$$

which is the SLLN.